Aspiration length and intensity as potential acoustic cues to [voice] in final position: evidence from Shughni

Yury Makarov

HSE University / Institute of Linguistics of the Russian Academy of Sciences
yurmak@iling-ran.ru

ABSTRACT

Final Shughni plosives tend to be released and aspirated, at least in citation forms. Data from 14 Shughni speakers showed that underlyingly voiced plosives have a shorter period of aspiration, and its intensity is weaker compared to that of voiceless ones. Although sometimes neutralized with regard to closure voicing, the contrast may be maintained by means of final aspiration length and intensity. Nevertheless, typological accounts do not usually treat these two characteristics as potential acoustic cues to underlying voicing in final position, and this study gives evidence that final aspiration length and intensity should be added to the list (cf. such cues as the duration of closure and preceding vowel, closure voicing, and F_1 frequency). Moreover, possible implications for languages with final aspiration (like Kashmiri) are discussed.

Keywords: neutralization, final devoicing, aspiration, acoustic cues, Iranian languages.

1. INTRODUCTION

1.1. Shughni language and speakers

Shughni is an Iranian language spoken by ca. 100,000 in the Pamir Mountains in Tajikistan and Afghanistan. It has several varieties, constituting the so-called Shughni-Rushani group. From now on, the variety spoken in Khorugh, the capital of the Gorno-Badakhshan Autonomous Region in Tajikistan, is described. Shughni is used primarily as a spoken language, though there are a number of poetry books and even novels. It is usually the case that Shughni speakers are bi- or even multilingual, having some knowledge of Tajik (official language) and Russian (lingua franca) [1].

1.2. Final laryngeal neutralization

A typical example of final laryngeal neutralization is final devoicing. It occurs when consonants of the voiced and voiceless series are realized without vocal fold vibration in final position, e.g., Russian /kod/ 'code' and /kot/ 'cat' pronounced as [kot]. Although neutralization may be complete in respect of closure voicing, acoustic cues to underlying voicing are likely to be present (thus making the neutralization process incomplete). For example, utterance-final postvocalic stops in English may have closure voicing, but there are other means of signalling the [voice] feature [2]. In fact, for American English, vowel duration is claimed to be a perceptually sufficient cue to underlying voicing in word-final stops, fricatives and clusters [3], [4]. On the other hand, in German, it was shown to be significant only in production [5].

Well-established acoustic cues to voicing in English are summarized in Table 1.

Cue	Voiced	Voiceless
preceding vowel	longer	shorter
closure	voiced shorter	voiceless longer
F_1 at the edges of adjacent vowels	lower	higher

Table 1: Acoustic cues to underlying voicing in utterance-final postvocalic stops in English (after [2]).

To be considered an acoustic cue, an acoustic characteristic must aid in the recognition of some feature of the segment [6], [7]. Until this perceptual significance is shown, this acoustic characteristic may be regarded as a potential cue, provided that it correlates with the feature in question. As for [voice], potential cues are the presence/absence of release (voiced stops are less often released) [8], [9], the intensity and spectral properties of release [10], and burst (expected differences are not specified) and aspiration (longer in underlyingly voiceless) duration [5], [7], [11] There is also a claim that 'any noise attendant upon consonant production will have greater intensity in the voiceless sound' [12], though few works are focused on considering the linguistic implications of this acoustic fact, regarding final stops in particular.

Final devoicing is not the only possible scenario of neutralization [13]. For instance, in Kashmiri (Indo-Aryan) the unaspirated and aspirated series are neutralized in favour of the latter, i.e., /wat/ 'way' and /kat^h/ 'story' are realized as [wat^h] and [kat^h] [14]. An interesting question is then what kind of cues (or at least acoustic correlates) are employed to maintain the underlying contrast when final aspiration occurs?

1.3. Final obstruents in Shughni

As recent studies have shown, plosives in Shughni are normally released in final position, and the preceding vowel is significantly longer before the underlyingly voiced consonant [15]. Final obstruents may be devoiced word-finally (F_0 present for less than half of the closure); it is less probable in careful speech. Interestingly, even if the final closure is fully voiced, the burst can be followed by a period of voiceless aspiration, see Fig. 1.

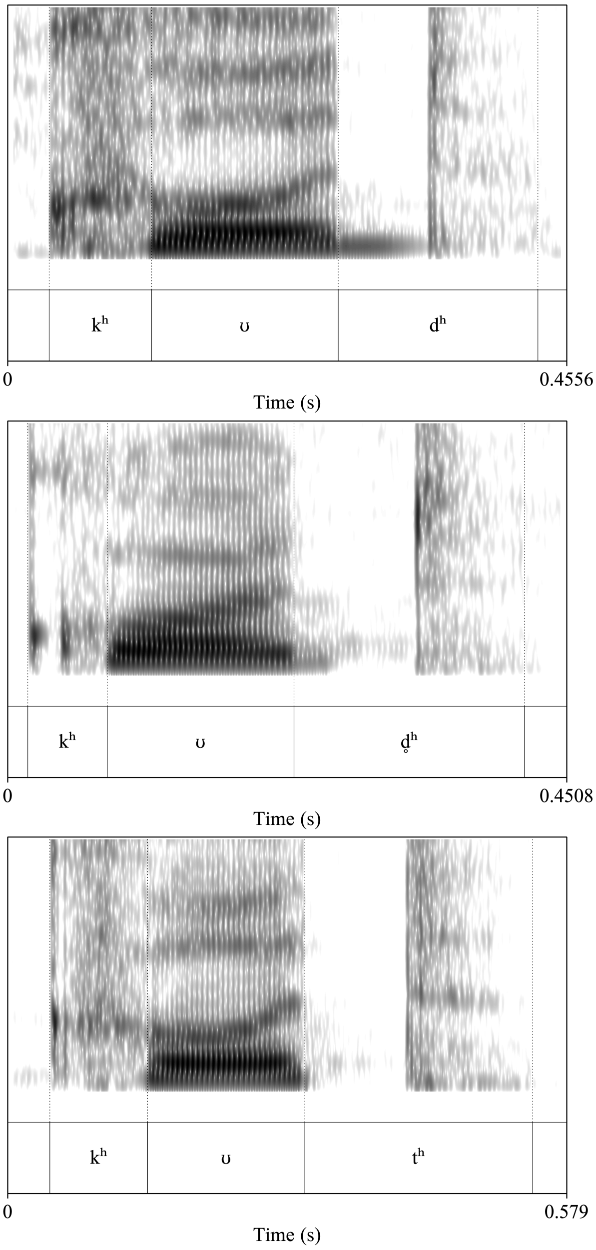

Figure 1: Dynamic spectrograms for /kʊd/ 'dog' (produced with a fully voiced and a devoiced final plosive) and /kʊt/ 'short.' Note voiceless aspiration in all cases. The view range is 0 to 5500 Hz.

The difference between /kʊd/ 'dog' and /kʊt/ 'short', though, impressionistically remains quite perceivable: the aspiration of final /t/ is 'stronger.'

1.4. Research purpose

In this paper, it is proposed that aspiration duration and intensity are the acoustic correlates of the underlying voicing contrast in word-final plosives in Shughni. The hypothesis is that underlyingly voiceless plosives have a longer and more intensive period of aspiration than voiced ones. The purpose of the next two sections is to test this hypothesis statistically. After that, implications for voiced/voiceless distinction and the typology of final laryngeal neutralization in phonetic theory are discussed.

2. DATA AND THEIR ANALYSIS

Data for this research were collected during fieldwork in Khorugh in 2022.

2.1. Speakers

14 native speakers of Shughni participated in the recording (mean age 28.9, SD 18.8; 12 females, 2 males). In accordance with Section 1.1, they all were bi- or multilingual. The speakers were not paid for participation, though received souvenirs.

2.2. Recording procedure and stimuli

The speaker and researcher sat in a quiet room. The researcher asked the speaker to translate a Russian word into Shughni. The stimuli were given along with fillers (they were Shughni words of the CVC structure ending in other obstruents and sonorants); the wordlist was obtained using *pamiri.online* [16]. If the speaker faced some trouble, an additional Russian synonym was given. Once the speaker suggested the intended translation, they were asked to repeat it three times with pauses between each of the utterances.

Shughni stimuli along with their English translations are given in Table 2.

Shughni	English
/bad/	bad
/bat/	embrace
/jed/	bridge
/jet/	open
/vud/	(it) was
/but/	footwear
/baːd/	then
/baːt/	a kind of national dish; bat
/kʊd/	dog
/kʊt/	short

Table 2: IPA transcriptions of the Shughni stimuli with their English translations.

A dynamic headset microphone (Shure WH20 XLR), connected to a digital recorder (Tascam DR-40x), was used to record utterances as 44.1 kHz 16-bit mono WAV files.

2.3. Data annotation

Only correctly pronounced words were annotated in Praat [17]. An utterance was considered mispronounced if there was no pause at its end or if the word was mumbled.

Every final plosive was segmented into closure, burst and aspiration; vowel duration was also annotated so as to be used as a reference point in relative measurements.

2.4. Data analysis

All of the annotated parts—namely vowel, final plosive's closure, burst and aspiration—were extracted from the recordings and represented as separate Sound objects in Praat. After that, duration and intensity were extracted using a Praat script. The `Get intensity (dB)` function was used in Praat for measuring intensity [18].

On average four pronunciations for a speaker were obtained: one when suggesting a Shughni translation and three when repeating in isolation. One of the speakers occasionally repeated the translation four times (resulting in 5 pronunciations in total); I decided to include all of them. The resulting spreadsheet with the data used in this study is available in [19].

In addition to the absolute values of the duration and intensity of aspiration, relative values were obtained using (1). Unlike absolute values, relative ones are expected to remain valid even when comparing two pronunciations with different intensity (due to dissimilar recording circumstances and/or manners of speaking): one with overall greater intensity and another with overall lesser one. While absolute measurements will likely show a significant difference, relative ones will not.

(1) $Relative = \frac{Absolute}{Vowel}$, where $Vowel$ is for the absolute duration or intensity of the preceding vowel.

I used linear mixed models from the *lme4* package [20] in R [21] to test the hypothesis from Section 1.4 statistically. The final plosive (/d/ or /t/) was considered the fixed effect, and the intercepts for word meaning, speaker and utterance number were included as random effects. The formula is shown in (2).

(2) $DurationORIntensity \sim FinalPlosive + (1 \mid WordMeaning) + (1 \mid Speaker) + (1 \mid UtteranceNumber)$

2.5. Limitations

The dataset has a few shortcomings. First, only one place of articulation (alveolar plosives) was taken into consideration. Second, careful speech was elicited, and, according to some sources [22], 'hyperarticulated speech may transform unreleased stops into released stops.' Third, there is a gender imbalance in favour of female speakers; as for the age, half of the speakers are 13 to 16; the other half is above 30 so at least two generations are represented.

Nevertheless, all these limitations, at least impressionistically, do not seem to be critical. Other places of articulation appear to have similar tendencies, and citation forms, while confined to specific contexts, are eligible subject matter.

3. RESULTS

Overall, the assumed presence of the [voice] feature in the final plosive significantly affected duration and intensity either in absolute or relative values. Table 3 contains the details.

All p-values (i.e., Pr(>|t|) values) are far below 0.01, though since they simply indicate that the estimates of the intercepts and differences between /t/ and /d/ are significantly different from 0, these data are of little interest for the current research purpose.

Fig. 2 depicts error bars for the reported measurements.

Response	Estimate	Standard error
Absolute Duration	0.044 s	0.004
Relative Duration	0.385	0.031
Absolute Intensity	2.877 dB	0.665
Relative Intensity	0.047	0.008

Table 3: Estimates of the difference between the expected values for /t/ and those for /d/ along with standard errors for each response obtained from the linear mixed models in Section 2.4.

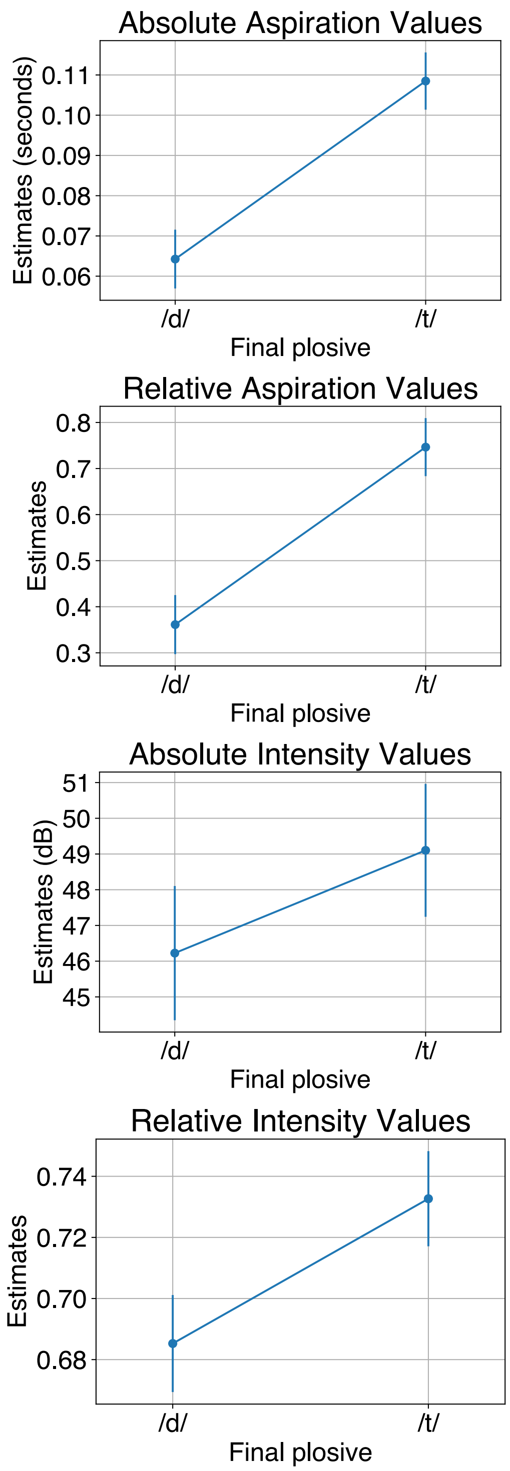

Figure 2: Error bars for each type of response.

4. DISCUSSION

The purpose of this study was to find out whether the duration and intensity of aspiration of final plosives in Shughni serve as acoustic correlates of underlying voicing and, hence, are potential acoustic cues to it. The hypothesis, based on both fieldwork observations and typological expectations, was that voiceless plosives would exhibit higher values either for duration or intensity compared to voiced ones.

The analysis of the data reveals that, regardless of the means of measurement (expressed in absolute or relative values), in final position, voiceless plosive /t/ has a longer period of aspiration with greater intensity. The difference between the predicted values of the absolute duration of aspiration in /t/ and /d/ is 44 ± 4 ms. It seems to be perceptible since such a duration is comparable to half of the vowel in /kʊt/ 'short.' Moreover, the just noticeable difference (JND) in discrimination of duration is estimated as a square root of the full duration of the stimulus (if this duration is lower than 100 ms) [23], [24] so a difference of 44 ms is quite discernible. The same is true for the difference of 2.877 ± 0.665 dB, with minimal values of the corresponding JND ranging from 0.3 to 1.0 dB [ibid.]. As for the relative values, their estimated differences are all significantly different from 0.

From a linguistic perspective, these results have three sides: one language-specific and two relevant to phonological typology.

The language-specific outcome is that Shughni final plosives (alveolar ones were tested, other places are assumed to have similar trends) may not be completely neutralized in final position. Although final devoicing sometimes occurs, underlyingly voiceless plosives have a longer period of aspiration of greater intensity in comparison to underlyingly voiced ones. This is a potential acoustic cue (among other candidates, see Section 1.3) and an additional study of Shughni plosives is needed to assess its role in perception. While the existing descriptions of Shughni do not agree on whether Shughni voiceless plosives are aspirated [25], [26], final aspiration supports adding [spread glottis] to their phonological specification.

As for the typological relevance, aspiration duration and intensity as acoustic correlates of underlying voicing have not been thoroughly studied (cf., for instance, [7], [22] and also Section 1.2 above). The Shughni data give evidence for adding them to the inventory of potential acoustic cues to [voice] in final position.[1]

Lastly, the current study provides support to the expectation that in a language with final aspiration (like Kashmiri above, see Section 1.2) the neutralization process may be incomplete due to the difference in aspiration duration and intensity. The similarity between the voiced unaspirated series (like

[1] This is also supported by the studies on final devoicing in German [11], [27], [28]; all of these, though, lack the instrumental measurement of intensity, and the difference in aspiration duration was generally less significant than found by my study. I thank Timofey Mets for drawing my attention to that.

that of Shughni) and the voiceless unaspirated one (like that of Kashmiri) is that neither of them is specified with the feature [spread glottis] (taken as the privative one, cf. 'laryngeal realism' [13], [29]). The Shughni pattern suggests that mere phonetic ('automatic') aspiration not reinforced with the underlying [spread glottis] feature is by default less prominent, especially in time and intensity domains.

5. ACKNOWLEDGEMENTS

I am indebted to Dr. George Moroz for his help with data analysis as well as to the Shughni speakers and the University of Central Asia, without whom this research would not have been possible. I am also grateful to the two anonymous reviewers, whose informative comments helped a lot, and Dr. Pavel Iosad, who pointed out the role of intensity in this case of incomplete neutralization. The publication was prepared within the framework of the Academic Fund Program at HSE University (grant № 22-00-034 'Field and Corpus Research on the Languages of the Shughni-Rushani Group').

6. REFERENCES

[1] D. (Joy) I. Edelman and L. R. Dodykhudoeva, 'Shughni', in *The Iranian languages*, G. Windfuhr, Ed., London and New York: Routledge, 2009, pp. 787–824.
[2] J. Kingston and R. L. Diehl, 'Phonetic Knowledge', *Language*, vol. 70, no. 3, pp. 419–454, 1994, doi: 10.1353/lan.1994.0023.
[3] L. J. Raphael, 'Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English', *J. Acoust. Soc. Am.*, vol. 51, no. 4B, pp. 1296–1303, Apr. 1972, doi: 10.1121/1.1912974.
[4] R. F. Port and J. Dalby, 'Consonant/vowel ratio as a cue for voicing in English', *Percept. Psychophys.*, vol. 32, no. 2, pp. 141–152, Mar. 1982, doi: 10.3758/BF03204273.
[5] T. B. Roettger, B. Winter, S. Grawunder, J. Kirby, and M. Grice, 'Assessing incomplete neutralization of final devoicing in German', *J. Phon.*, vol. 43, pp. 11–25, Mar. 2014, doi: 10.1016/j.wocn.2014.01.002.
[6] R. L. Trask, *A dictionary of phonetics and phonology*, Reprinted. London; New York: Routledge, 2006.
[7] D. Steriade, 'Phonetics in Phonology: The Case of Laryngeal Neutralization', 1997.
[8] W. S.-Y. Wang, 'Transition and Release as Perceptual Cues for Final Plosives', *J. Speech Hear. Res.*, vol. 2, no. 1, pp. 66–73, Mar. 1959, doi: 10.1044/jshr.0201.66.
[9] J. E. Flege and J. Hillenbrand, 'A differential effect of release bursts on the stop voicing judgments of native French and English listeners', *J. Phon.*, vol. 15, no. 2, pp. 203–208, Apr. 1987, doi: 10.1016/S0095-4470(19)30548-0.
[10] M. Halle, G. W. Hughes, and J. -P. A. Radley, 'Acoustic Properties of Stop Consonants', *J. Acoust. Soc. Am.*, vol. 29, no. 1, pp. 107–116, Jan. 1957, doi: 10.1121/1.1908634.
[11] R. F. Port and M. L. O'Dell, 'Neutralization of syllable-final voicing in German', *J. Phon.*, vol. 13, no. 4, pp. 455–471, Oct. 1985, doi: 10.1016/S0095-4470(19)30797-1.
[12] D. B. Fry, *The physics of speech*. in Cambridge textbooks in linguistics. Cambridge: Cambridge University Press, 1996.
[13] G. K. Iverson and J. C. Salmons, 'Final Devoicing and Final Laryngeal Neutralization', in *The Blackwell Companion to Phonology*, M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice, Eds., Oxford, UK: John Wiley & Sons, Ltd, 2011, pp. 1–22. doi: 10.1002/9781444335262.wbctp0069.
[14] S. M. Syeed, 'The Himalayan way of breathing the last: On the neutralizing status of the word-final aspiration in Kashmiri', *East. Anthropol.*, no. 31, pp. 531–541, 1978.
[15] Y. Makarov, 'Plosives in Shughni', HSE University, Moscow, 2022. [Online]. Available: https://www.researchgate.net/publication/361584042_Realizacia_vzryvnyh_soglasnyh_v_sugnanskom_azyke_Plosives_in_Shughni
[16] Y. Makarov, M. Melenchenko, and D. Novokshanov, 'Digital Resources for the Shughni Language', in *Proceedings of The Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia within the 13th Language Resources and Evaluation Conference*, Marseille, France: European Language Resources Association, Jun. 2022, pp. 61–64. [Online]. Available: https://aclanthology.org/2022.eurali-1.9
[17] P. Boersma and D. Weenik, 'Praat: doing phonetics by computer [Computer program]'. Apr. 15, 2021. [Online]. Available: http://www.praat.org/
[18] ppgb, 'Sound: Get intensity (dB)', *Praat Manual*, Jan. 29, 2007. https://www.fon.hum.uva.nl/praat/manual/Sound__Get_intensity__dB_.html
[19] Y. Makarov, 'Raw data (measurements)'. Unpublished, 2023. doi: 10.13140/RG.2.2.23847.62888.
[20] D. Bates, M. Mächler, B. Bolker, and S. Walker, 'Fitting Linear Mixed-Effects Models Using lme4', *J. Stat. Softw.*, vol. 67, no. 1, 2015, doi: 10.18637/jss.v067.i01.
[21] R Core Team, 'R: A Language and Environment for Statistical Computing'. R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/
[22] J. Blevins, *Evolutionary Phonology: The emergence of sound patterns*. New York: Cambridge University Press, 2004.
[23] K. N. Stevens, *Acoustic phonetics*. in Current studies in linguistics series, no. 30. Cambridge, Mass.: MIT Press, 2000.
[24] E. Koffi, 'A Comprehenisve Review of the Acoustic Correlate of Duration and Its Linguistic Implications', *Linguist. Portf.*, vol. 10, 2021, [Online]. Available: https://repository.stcloudstate.edu/stcloud_ling/vol10/iss1/2
[25] K. Olson, *Shughni Phonology Statement*. SIL International, 2017.
[26] D. Karamshoev, *Badzhuvskij dialekt shugnanskogo jazyka [Bajuwi dialect of Shughni]*. Dushanbe: Izdatel'stvo AN Tadzhykskoj SSR, 1963.
[27] M. Fourakis and G. K. Iverson, 'On the "Incomplete Neutralization" of German Final Obstruents', *Phonetica*, vol. 41, no. 3, pp. 140–149, May 1984, doi: 10.1159/000261720.
[28] M. O'Dell and R. Port, 'Discrimination of word-final voicing in German', *J. Acoust. Soc. Am.*, vol. 73, no. S1, pp. S31–S31, May 1983, doi: 10.1121/1.2020331.
[29] P. Honeybone, 'Germanic Obstruent Lenition; some mutual implications of theoretical and historical phonology', PhD Thesis, University of Newcastle upon Tyne, Newcastle upon Tyne, 2002.