Yury Makarov^{1,2}, Mark Gibson³, and Bert Vaux¹ ¹University of Cambridge; ²Institute of Linguistics, RAS; ³Universidad de Navarra

Reconstructing the perception of [± voice] in Shughni: is vowel quality enough?

Recent research has shown that in Shughni, a minority Iranian language spoken in Tajikistan and Afghanistan, final laryngeal neutralisation is incomplete, and perceptual cues for voicing are not predictable from the corresponding production cues [1], [2]. Most notably, final plosives differ significantly in aspiration intensity and duration (e.g., /t/ is 44 ± 4 ms longer / 2.877 \pm 0.665 dB more intense than /d/ [1]). However, a perception experiment revealed no effect of final aspiration on the discernibility of minimal pairs such as /kod/ 'dog' vs. /kot/ 'short.' This misalignment makes it unclear why such significant differences between the two obstruent series are present in production without having obvious use in perception.

The goals of the current study included reverifying the findings concerning final aspiration using a different experimental design and investigating the role of other production cues in the perception of final obstruent voicing. Like the previous study, we used a forced-choice paradigm. However, in our perception test, we asked 16 Shughni speakers to choose whether they heard /kod/ 'dog' or /kot/ 'short' when 74 modified stimuli × 3 repetitions (absent from the previous test) were played. Furthermore, instead of using only two values for stimuli corresponding to a particular cue (e.g., long vs. short closure duration), we used temporal/spectral means to calculate hybrid values for the tested acoustic parameters. These neutralised hybrid values were subsequently increased/decreased in the manipulated stimuli in intervals of 7.5%, 15%, and 22.5%, yielding six audio files for each acoustic parameter. This approach allowed us to establish not only the presence/absence of the weight of a cue in perception but also the threshold after which the cue gets enabled. Lastly, to control for the potential influence of vowel quality (not done in the previous test), we modified the natural recordings of both /kod/ and /kot/.

The effects of incremental increases/decreases of the acoustic parameters *vowel duration* (*VD*), *closure duration* (*CD*), *aspiration duration* (*AD*) and *aspiration intensity* (*AI*) on the final obstruent voice discrimination were tested using generalised linear mixed models with the "lme4" package [3] in RStudio. We found that neither AD nor AI is a reliable acoustic cue for voicing in Shughni, see Fig. 1. However, longer AD led to a decrease in /d/-interpretations as would be expected if longer AD were indeed associated with voicelessness.



Figure 1. Accuracy percentage for aspiration duration and intensity.

Overall, there was a main effect for original vowel (preceding /d/ or /t/) across all studied parameters, which might be considered a sign that vowel timbre is the main acoustic cue for voicing. The situation is complicated by the fact that the results for VD and CD, where the effect in question was also partly observed, indicated that VD and CD significantly influence the perception of stimuli with the vowel originally preceding /d/ but not /t/ (see Fig. 2).



Figure 2. Accuracy percentage for vowel and closure duration. For /kod/-based stimuli only, longer VD resulted in more [+ voice] interpretations as well as shorter CD.

Based on our results, we propose the following hierarchy of acoustic cues for final obstruent voicing in Shughni (drawing on /kod/-based stimuli):

More significant

Less significant

vowel duration, closure duration < vowel quality < aspiration intensity, aspiration duration

Since the accuracy rate for /kot/-based stimuli remained high (>80%) regardless of changes in other acoustic parameters, we have to assume that the vowel quality typical of this [– voice] context overrides other cues. This poses a theoretical problem: a single phonological feature ([\pm voice] in our study) is expected to have a uniform set of acoustic cues. In Shughni, however, we find that once [– voice] is established based on vowel quality, the other cues are disregarded, and only if vowel quality implies [+ voice], VD and CD are taken into account. In a normal situation, we would expect one cue (vowel quality) to have constant weight in computing the value of a feature, which would be the case if VD and CD overrode vowel quality or vice versa.

References

- [1] Y. Makarov, 'Aspiration length and intensity as potential acoustic cues to [voice] in final position: evidence from Shughni', in *Proceedings of the 20th International Congress of Phonetic Sciences*, R. Skarnitzl and J. Volín, Eds., Prague: GUARANT International spol. s r.o., 2023, pp. 2298–2302. [Online]. Available: https://guarant.cz/icphs2023/582.pdf
- [2] Y. Makarov, 'The phonetics and phonology of Shughni', MPhil dissertation, University of Cambridge, Cambridge, United Kingdom, 2024.
- [3] D. Bates, M. Mächler, B. Bolker, and S. Walker, 'Fitting Linear Mixed-Effects Models Using lme4', *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015, doi: 10.18637/jss.v067.i01.