A Crosslinguistic Analysis of Intrinsic Vowel Duration

Introduction. Across languages, it has commonly been observed that high vowels tend to have a shorter duration than low vowels (Maddieson, 1996; Lehiste, 1970). The observed consistency could arise from biomechanical factors: assuming that the duration target for all vowels is uniform, the motor movements involved in a low vowel production (e.g., jaw lowering) could require more time than in a high vowel production. Alternatively, speakers could be "in control" of such differences, and might even enhance the biomechanical effect using this secondary cue (Solé and Ohala, 2010).

The current study presents an empirical investigation of intrinsic vowel duration in 30 languages from nine language families, as well as an analysis of the nature of the relationship. We investigate the following hypotheses: If the effect is biomechanical, high vowels should be universally shorter than low vowels to a consistent degree across speakers and languages. Any deviation would suggest that speaker control is a viable option (cf., Solé and Ohala, 2010). If speaker control is apparent, two broad possibilities emerge: first, speakers could theoretically specify the high and low vowel duration targets in any way possible, resulting in no observable structure in the relationship. Alternatively, speakers could conform to their language population, revealing structure in the relationship on a language-specific level. However, additional sub-constraints could further constrain the nature of the relationship within the population. Namely, languages could prefer near-uniformity in the specification, with minimal deviation from the expected biomechanical effect.

Data and Processing. Speech data was sourced from the Mozilla Common Voice Dataset (Ardila et al., 2020), with phone- and word-level alignments from the VoxCommunis Corpus (Ahn and Chodroff, 2022). Languages were considered for analysis if they had a high front and low vowel category, as well as more than two hours of validated data for the forced alignment acoustic model. The final analysis included data from 30 languages from nine language families. The number of analyzed hours ranged from 2 to 339 with a median of 43 hours per language.

Vowel duration was measured for high front vowels [i ɪ] and low vowels [a ɑ]. A high front and low vowel category was created for each language by collapsing or removing tone, nasalization, tense/lax distinctions on a case-by-case basis. Long and short vowel contrasts were separated into distinct classes for analysis, resulting in 36 high–low contrasts. Duration was rate-normalized for each vowel by dividing the duration by the speaking rate, where speaking rate was defined as the number of phones per utterance and the total duration per utterance, excluding the target vowel (Tilsen and Tiede, 2023). Following outlier exclusion, a median of 35210 vowels per language were available for analysis.

Analysis and Results. For each language, we estimated the effect of phonological vowel height on standardized vowel duration using a linear mixed-effects model with a random speaker intercept. Across languages, we observed substantial differences in the effect of vowel height on vowel duration (see Figure 1). The effect of vowel height was significant and in the expected direction for 27 out of 36 classes, though the magnitude ranged considerably. For 7 classes, the effect was significant and in the opposite direction, and for two classes, the effect did not reach significance (Bulgarian and Catalan). The fit between height and duration had a moderate to strong R^2 value for all languages with the exception of Marathi and Hindi (see Figure 2).

Conclusion. The crosslinguistic trend conforms with the expected direction that would be predicted from the biomechanical effect. Nevertheless, languages do deviate from this norm and to differing degrees, suggesting speaker-specific control. Overall, languages appear to allow for unique duration targets for /i/ and /a/ and with a strong bias towards the biomechanical difference; however, speakers are subject to strong constraints that promote predictability within the language (a community-level bias) and to a lesser degree, minimize duration differences.

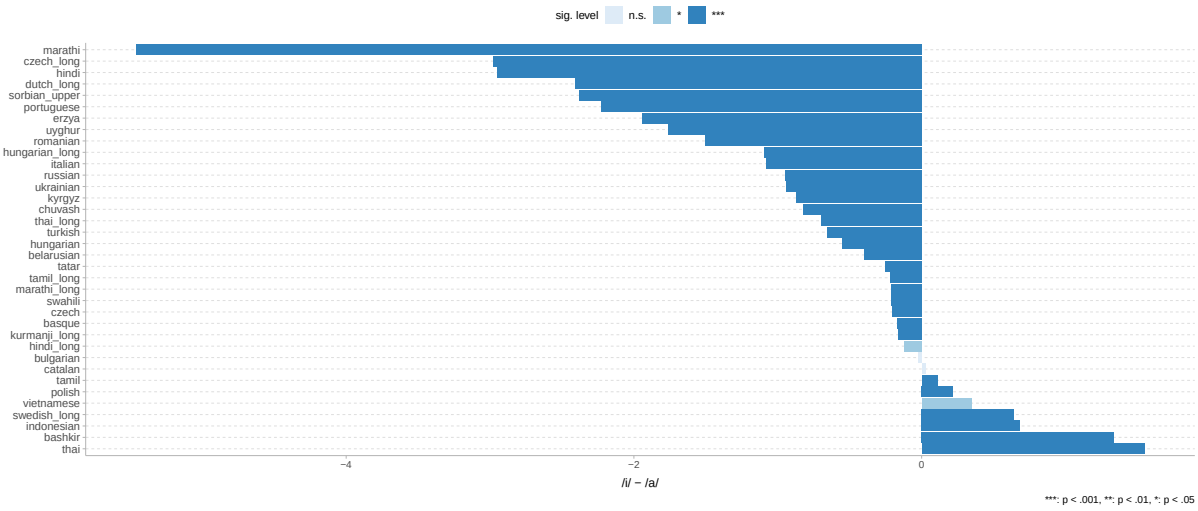

Figure 1: The estimated difference in standardized duration between /i/ and /a/ across languages. Linear mixed-effect models were fit to each language.

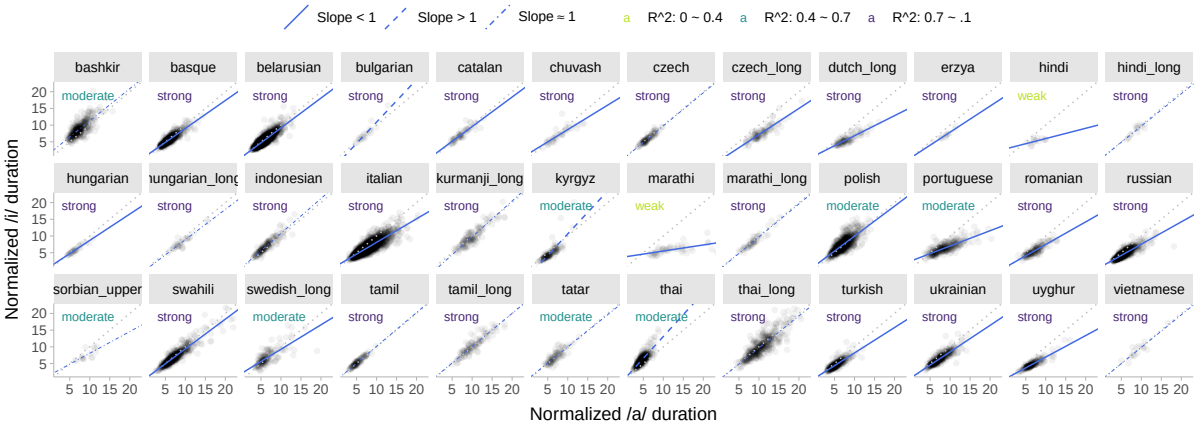

Figure 2: The linear regression lines fit to each language. Each data point represents a speaker.

References

Ahn, E. P. and Chodroff, E. (2022). VoxCommunis: A Corpus for Cross-linguistic Phonetic Analysis. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pages 5286–5294.

Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., and Weber, G. (2020). Common Voice: A Massively-Multilingual Speech Corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 11–16.

Lehiste, I. (1970). *Suprasegmentals*. MIT Press, Cambridge, MA.

Maddieson, I. (1996). Phonetic universals. *UCLA Working Papers in Phonetics*, pages 160–178.

Solé, M.-J. and Ohala, J. J. (2010). What is and what is not under the control of the speaker: intrinsic vowel duration. *Papers in Laboratory Phonology 10*, (January):607–655.

Tilsen, S. and Tiede, M. (2023). Parameters of unit-based measures of speech rate. *Speech Communication*, 150:73–97.