Phoneme Inventory Effects on Speaker Specificity in /n/: A Cross-Linguistic Study

Annie Baker[1], *Eleanor Chodroff*[1]
[1]*Department of Computational Linguistics, University of Zurich, Switzerland*
annie.baker@uzh.ch eleanor.chodroff@uzh.ch

Introduction

Nasal segments have been shown to exhibit relatively high levels of speaker specificity across numerous languages including Standard Dutch (Smorenburg & Heeren, 2021), Japanese (Amino & Arai, 2009), Standard Southern British English and Leeds English (Kavanagh, 2012), as well as Spanish and Catalan (Gavalda & Cicres, 2025). The present study investigates whether the phoneme inventory influences the extent of speaker-specific variation by examining the degree of speaker specificity in the alveolar nasal /n/ across languages with considerably different nasal inventories. We compare Abkhaz, whose phoneme inventory includes two nasals /n/ and /m/, with Tamil, which has six (/n, m, n̪, ɲ, ɳ, ŋ/). We hypothesise that speakers of languages with few nasal contrasts may have greater freedom for individual variation, being that they are less constrained by the need to preserve a large number of linguistic contrasts, in this case along the place of articulation dimension. As nasal place contrasts are cued acoustically by spectral properties arising from vocal-tract filtering and antiresonance structure (Stevens, 2000), the analysis focuses on acoustic measures which may encode both phonological and speaker-specific information.

Methods

Speech data for this study come from the Mozilla Common Voice corpus (Ardila et al., 2020) version 17, with corresponding time-aligned TextGrids from the VoxCommunis Corpus (Ahn & Chodroff, 2022; Zhang et al., 2025a; Zhang et al., 2025b). Grapheme-to-phoneme conversion was implemented using the XPF Corpus for Abkhaz (Cohen Priva et al., 2021) and Epitran for Tamil (Mortensen et al., 2018). For each language, 200 speakers with at least 10 tokens of /n/ were chosen at random from the broader corpus. The following acoustic properties were then measured in hertz from the middle 20 ms of each /n/ segment following multitaper spectral analysis (Shadle et al., 2012; Chodroff & Wilson, 2014): centre of gravity (COG), spectral standard deviation (SD) (square root of spectral variance) and the spectral minimum between the estimated F1 and F2 values, representing the first anti-formant (Stevens, 2000; Kavanagh, 2012; Gavalda & Cicres, 2025). F1 and F2 of the nasal were estimated using Praat with the utterance-specific parameters for formant extraction based on the mean f0 of the inputted utterance. The Jensen-Shannon divergence (JSD) was then calculated (Lin, 1991) between each pair of speaker distributions for each measure. JSD provides a symmetric distance metric between two distributions, with higher JSD values representing greater distance between the distributions, and therefore higher speaker specificity.

Results and Discussion

For all three acoustic properties (COG, SD, spectral minimum), a linear mixed effects model was fitted to predict JSD, with the fixed effect of language and a random by-speaker intercept. The preliminary results show that, contrary to expectations, the JSD values were significantly higher for Tamil than for Abkhaz across all three acoustic properties, with the most marked difference being JSD for COG (see Figure 1). The Tamil data also showed higher variance for all JSD measures, as indicated by the standard deviations. These findings suggest that languages differ in speaker specificity for nasal segments, but phoneme inventory size may not reliably predict this variation, at least along these dimensions. Next steps for this project include further investigation into establishing relevant acoustic dimensions for place of articulation contrasts, and analysis of further spectral measures such as amplitude measures.

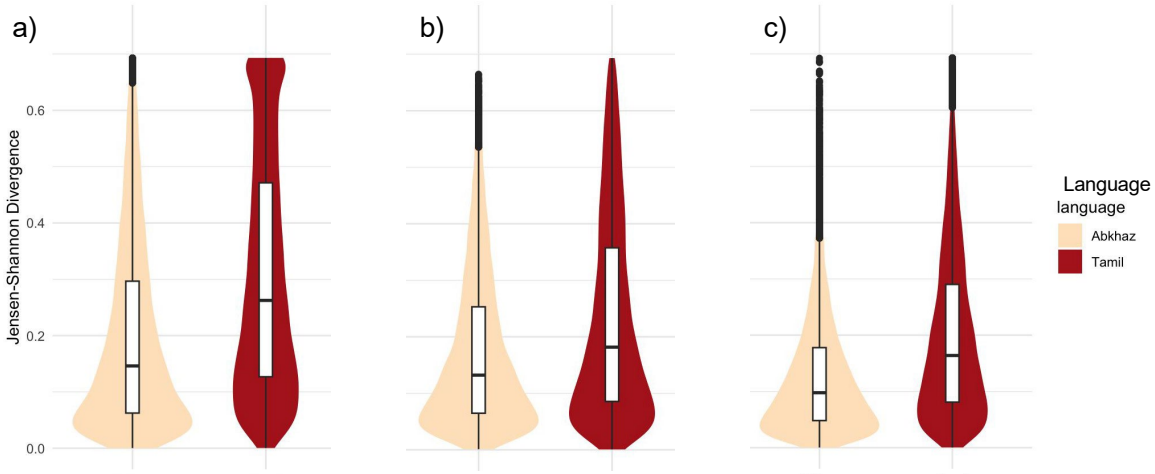

Figure 1. Violin and boxplots showing the Jensen-Shannon divergence values for a) centre of gravity,b) spectral standard deviation, and c) spectral minimum for Abkhaz and Tamil.

References

Ahn, E., & Chodroff, E. (2022). *VoxCommunis: A corpus for cross-linguistic phonetic analysis*. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 5286–5294).

Amino, K., & Arai, T. (2009). Speaker-dependent characteristics of the nasals. *Forensic Science International, 185*(1–3), 21–28.

Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2020). Common Voice: A massively-multilingual speech corpus. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 4218-4222).

Chodroff, E., & Wilson, C. (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.https://doi.org/10.1121/1.4896470

Cohen Priva, Uriel and Strand, Emily and Yang, Shiying and Mizgerd, William and Creighton, Abigail and Bai, Justin and Mathew, Rebecca and Shao, Allison and Schuster, Jordan and Wiepert, Daniela. (2021). The Cross-linguistic Phonological Frequencies (XPF) Corpus

Gavaldà, N., & Cicres, J. (2025). An exploration of the language independence and speaker-specificity of /m/ and /n/ in speakers of Catalan, Spanish, and English. *International Journal of Bilingualism*. Advance online publication. https://doi.org/10.1177/13670069251380015

Kavanagh, C. (2012). *New consonantal acoustic parameters for forensic speaker comparison* [Doctoral dissertation, University of York].

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory, 37*(1), 145–151. https://doi.org/10.1109/18.61115

Mortensen, D. R., Dalmia, S., & Littell, P. (2018, May). Epitran: Precision G2P for many languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Shadle, C. H., Cohn, A., Fougeron, C., & Huffman, M. (2012). Acoustics and aerodynamics of fricatives. In A.C. Cohn, C. Fougeron, & M. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 511–526). Oxford University Press.

Smorenburg, B., & Heeren, W. (2021). Forensic value of acoustic-phonetic features from Standard Dutch nasals and fricatives. In *Speaker Individuality in Phonetics and Speech Sciences: Speech Technology and Forensic Applications – Book of Abstracts* (pp. 46–47).

Zhang, M., Ahn, E., Baker, A., & Chodroff, E. (2025, February). *VoxCommunis Corpus*. OSF Repository. https://doi.org/10.17605/OSF.IO/T957V

Zhang, M., Farhadipour, A., Baker, A., Ma, J., Pricop, B., & Chodroff, E. (2025). Quantifying and reducing speaker heterogeneity within the Common Voice corpus for phonetic analysis. In *Proceedings of Interspeech 2025*, 3933-3937, doi: 10.21437/Interspeech.2025-2027