ACOUSTIC PROPERTIES OF THE FOUR-WAY LARYNGEAL CONTRAST IN BENGALI INFANT DIRECTED SPEECH

Jahnavi Narkar

University of California, Los Angeles
jnarkar@ucla.edu

ABSTRACT

This paper analyzes the acoustic properties of the four-way laryngeal contrast in Infant Directed Speech (IDS), comparing it with Adult Directed Speech (ADS). Since IDS has been demonstrated to be hyper-articulated, its acoustic manifestation is expected to contain enhanced cues to the four-way contrast. I investigated VOT, PVI and H1*–H2* of prevocalic stops and affricates in the speech of 10 Bangladeshi Bengali speakers. The results revealed an asymmetry in IDS – the VOT of voiced-unaspirated and voiced-aspirated segments was longer in IDS, but there was no difference in the VOT of voiceless-aspirated and voiceless-unaspirated segments between the two registers; PVI of voiced-aspirated segments, but not that of voiceless-aspirated segments, was longer in IDS. Finally, H1*–H2* did not differ by register, although the expected effect of category was found. Viewed against studies on similar languages, these results suggest that language-specific phonetic grammars modulate the acoustic manifestation of IDS.

Keywords: Infant directed speech, laryngeal contrast, acoustic cues, VOT

1. INTRODUCTION

Eastern Bengali (Bangladesh), also known as Bangla (henceforth referred to simply as *Bengali*), employs a four-way system of laryngeal contrasts that utilizes both voicing and aspiration. This four-way contrast is illustrated in Table 1 where 'T' denotes voiceless-unaspirated stops and affricates; 'Th' denotes voiceless-aspirated stops and affricates; 'D' denotes voiced-unaspirated stops and affricates; 'Dh' denotes voiced-aspirated stops and affricates.

Cross-linguistically, voicing contrasts have been typically characterized using voice onset time (VOT) which has been proposed to be the primary cue to voicing in languages with two or three-way distinctions [1]. However, VOT is not sufficient to distinguish the four-way contrast of languages like Bengali [2]. In addition to VOT, several other acoustic cues have been found to contribute to the four-way contrast, such as onset f0, burst frequency and measures of spectral tilt [3, 4, 5, 6].

Category	Segments	Example
T	[p, t̪, tɕ, ʈ, k]	[t̪ana] (*drawn*)
Th	[pʰ, t̪ʰ, tɕʰ, ʈʰ, kʰ]	[t̪ʰana] (*police station*)
D	[b, d̪, dʑ, ɖ, ɡ]	[d̪ana] (*grain*)
Dh	[bʱ, d̪ʱ, dʑʱ, ɖʱ, ɡʱ]	[d̪ʱana] (*paddy*)

Table 1: The Bengali four-way contrast.

In this paper, I tease apart the primary cues to this contrast in Bengali by comparing the acoustic properties of infant-directed speech (IDS) and adult-directed speech (ADS). IDS has been thought to be hyper-articulated to facilitate phonetic learning [7]. This is generally true for vowels which show greater separation between categories, but the results on consonants are mixed. Stops in languages that have a two-way voicing contrast have been found to be hyper-articulated [8, 9], hypo-articulated [10] and no different from each other [11] in terms of VOT. In these cases, where there is a single primary cue to voicing, hyper-articulation may involve lengthening the VOT of both stop categories such that the relative separation between them remains comparable. Conversely, if aspirated stops are produced with longer VOT, and unaspirated stops with comparable or shorter VOT, the two categories are more separated. I investigate the former case whereby VOT is lengthened, but it must be pointed out that this only involves enhancing the primary cue associated with the contrast, not necessarily increasing separation between categories.

The only study on IDS in a language with the four-way contrast is [12] which found that Nepali stops are hypo-articulated in IDS in terms of VOT. In addition to VOT, I measured two acoustic cues that have been proposed to be adequate for distinguishing this complex contrast – PVI (pre-vocalic interval) and H1*–H2*. Speakers are expected to enhance primary cues in hyper-articulated registers like IDS.

Therefore, if they are primary, VOT and PVI, which are durational cues, are expected to be lengthened in IDS, and H1*–H2*, which is a measure of breathy voice, is expected to be greater in IDS. In addition to providing a description of the acoustic properties of IDS, the results also have a bearing on the realization and representation of the four-way contrast, and on the role of IDS in phonetic acquisition.

2. METHOD

2.1. Materials

Data collected by [13] for their investigation of the intonational phonology of Bengali IDS were used in this study. They recorded IDS and ADS speech of 10 native speakers of Bengali reading the "North Wind and Sun" fable translated into Bengali, from [14]. For the ADS speech, subjects were asked to "read (the fable) at a comfortable pace." For the simulated IDS speech, subjects were asked to read the same passage as if speaking to their 4-5 month-old infant. Each speaker produced the IDS and ADS version three times each.

Target segments from the speech samples were coded as [p, b, b$^{\text{ɦ}}$, t̪, t̪$^{\text{h}}$, d̪, d̪$^{\text{ɦ}}$, tɕ, tɕ$^{\text{h}}$, dʑ, ʈ, k, k$^{\text{h}}$, g]. Measures for /p$^{\text{h}}$/ were not extracted since the speakers consistently produced these as [f]. There were no instances of /ʈ$^{\text{h}}$, ɖ, ɖ$^{\text{ɦ}}$, g$^{\text{ɦ}}$, dʑ$^{\text{ɦ}}$/ in the fable. Each pre-vocalic segment of interest was included. Since VOT is closely related to speech rate, both global speech rate (at the utterance level) and local speech rate (local to the target phone) were checked. Both rates were significantly slower in IDS. Therefore, the two temporal cues, VOT and PVI, are expected to be longer in IDS.

2.2. Acoustic Analyses

2.2.1. VOT

Category	VOT	PVI	H1*–H2*
T	2999	2999	1804
Th	518	518	403
D	1429	1429	1410
Dh	453	453	435

Table 2: Tokens included in the analyses.

VOT was measured in Praat [15] as the temporal interval between the beginning of the release burst and the onset of quasi-periodicity. For voiced stops and affricates (voicing lead), it was measured as the duration between the onset of voicing and the stop burst. For voiceless stops (voicing lag), VOT was measured as the duration between the burst and the onset of voicing. Tokens that were erroneous or had noise or interruptions were excluded. Table 2 shows the total number of tokens included in the final analyses.

2.2.2. Pre-vocalic interval (PVI)

Following [5], I measured PVI in Praat [15] as the temporal interval between the beginning of the release burst and the onset of breathiness in the following vowel as indicated by the clear onset of a dark F2 in the spectrogram or an increase in amplitude in the waveform. Notice that since PVI is a durational measure of the breathy portion of the stops, the crucial categories are the two aspirated ones – Dh and Th.

2.2.3. H1–H2**

The difference in amplitudes of the first and second harmonics, corrected for formant frequencies and bandwidths, H1*–H2*, was measured in Voicesauce [16] at the onset of the following vowel as previous studies have shown that difference in breathiness rapidly declines after the onset [5, 6]. The affricates were excluded from the analysis of H1*–H2* as there were no instances of breathy-voiced [dʑ$^{\text{ɦ}}$], and [tɕ], [tɕ$^{\text{h}}$] and [dʑ] were expected to have greater H1*–*H2* from the fricative portion following the stop burst. Therefore, the contribution of the breathy portion following voiced-aspirated stops would be obscured due to the greater H1*–H2* in the three other laryngeal categories driven by the affricates. To account for differences between speakers, H1*–H2* was z-scored by speaker to reduce variation. Tokens with absolute z-scores greater than 3 were taken to be outliers and excluded from the analysis (see Table 2). While the number of tokens is not distributed evenly across the laryngeal categories, they do reflect the frequency of segments belonging to the four categories in the lexicon.

2.3. Statistical Analyses

All statistical analyses were conducted using mixed effects regression models using the `lmer` function in the `lme4` package [17] in R [18]. To compare the effect of register on VOT, separate linear mixed effects models were run for each category. Each optimal model was obtained by applying the `step` function, which performs backward elimination of non-significant effects, to a fully specified model. This fully specified model contained all the fixed factors, their interactions, and random intercepts for speaker and word. Since VOT is known to vary with place of articulation [19] and prosodic position [20], the models of VOT and PVI included these as fixed factors in addition to register. Since there was no reason to expect an effect of place or position on breathiness, these factors were not included in the H1*–H2* model.

3. RESULTS

3.1. VOT

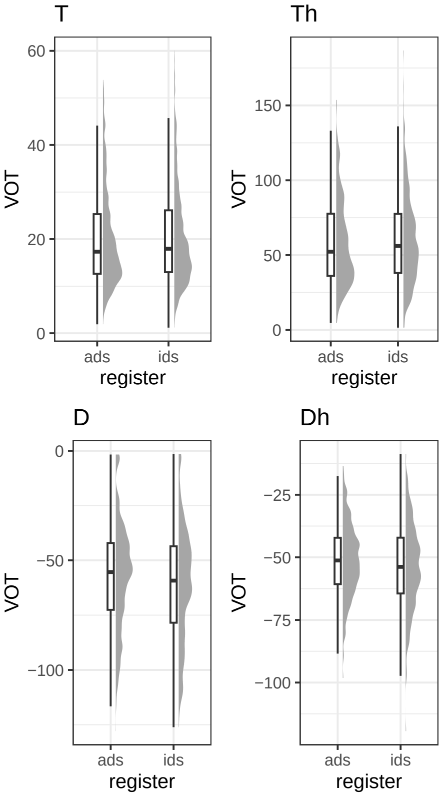

Figure 1: VOT by register.

Fig. 1 shows the VOT for each laryngeal category by register. Note that since the VOT of voiced stops is negative, for the Dh and D categories, greater absolute values of VOT (lower on the y-axis) correspond to longer durations. For the T category, there was no effect of register on VOT, which is in line with studies on English and Spanish which found that voiceless-unaspirated stops were not hyper-articulated in IDS even when voiceless-aspirated and prevoiced stops were [9]. Additionally, the expected result of place was found – [tɕ] had significantly longer VOT than [p]($\beta = -21.85, t(159) = -18.99, p < 0.001$), [t̪]($\beta = -24.06, t(85) = -19.18, p < .001$), [ʈ]($\beta = -24.93, t(320) = -20.26, p < .001$) and [k] ($\beta = -7.58, t(66) = -5.38, p < .001$).

For the Th category, there was no effect of register on VOT, contra [8, 9, 12]. However, the expected effects of place and word position were found – [tɕ$^{\text{h}}$] had significantly longer VOT than [k$^{\text{h}}$] ($\beta = -32.71, t(11) = -9.71, p < .001$) and [t̪$^{\text{h}}$] ($\beta = -49.7, t(10) = -14.22, p < .001$); VOT was longer in word-initial position than in word-medial position ($\beta = -8.44, t(12) = -2.48, p < .05$). Thus, this model shows that voiceless-aspirated stops in Bengali IDS are neither hyper- nor hypo-articulated.

For the D category, the optimal mixed effects model found a significant effect of register on VOT ($\beta = -4, t(1232) = -3.94, p < .001$) such that voiced stops in IDS had significantly longer negative VOT compared to ADS. There was no effect of place, but the expected effect of prosodic position on VOT was found ($\beta = 20.97, t(22) = 2.57, p < .05$). Finally, for the Dh category, there were significant effects of register ($\beta = -2.75, t(424) = -2, p < .05$) and place ($\beta = 22.73, t(6) = 3.4, p < .05$), but no effect of prosodic position. Thus, both voiced-unaspirated and voiced-aspirated stops and affricates in Bengali IDS were hyper-articulated in terms of VOT compared to ADS. The effect of prosodic position on the VOT of Th and D stops also serves to extend results from previous studies investigating the effect of prosodic position on VOT in other languages [20] to Bengali.

3.2. Pre-vocalic interval (PVI)

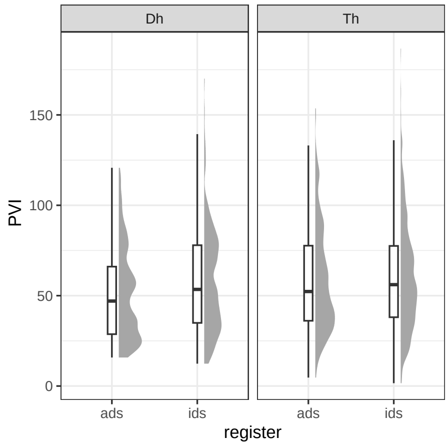

Figure 2: PVI by register.

Fig. 2 shows the PVI of the Dh and Th categories by register. For Dh, the linear mixed effects model showed a significant effect of register ($\beta = 5.66, t(398) = 2.67, p < .01$) – stops in IDS had longer PVI than in ADS. This implies that, like VOT, the aspiration portion of voiced aspirated stops is also exaggerated via lengthening in IDS. By contrast, the PVI of the Th segments was not affected by register. Thus, voiced-aspirated stops in IDS were hyper-articulated in terms of PVI while voiceless-aspirated stops were not hyper-articulated and had comparable PVI in both registers.

3.3. H1*–H2*

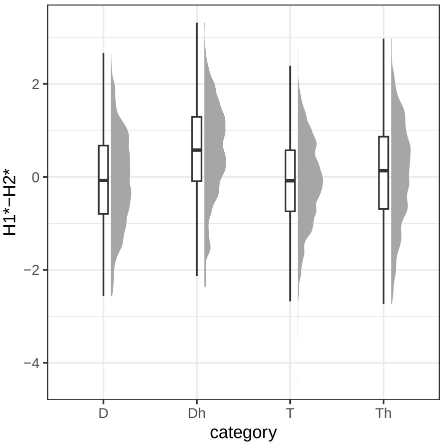

Figure 3: H1*–H2* by laryngeal category.

The linear mixed effects model did not find an effect of register on H1*–H2*. That is, IDS was not found to be breathier than ADS for any category. Since no effect of register was found, the data were combined into a single model to check for the effect of category on H1*–H2*. Fig. 3 shows this H1*–H2* by category with the IDS and ADS combined. This figure shows the expected effect of category – Dh is breathier than the rest of the laryngeal categories, and Th is slightly breathier than D and T. The full mixed effects model confirmed this. The only significant effect was that of category – Dh ($\beta = .93, t(118) = 7.76, p < .001$) and Th ($\beta = .17, t(482) = 2.38, p < .05$) were significantly breathier than T. H1*–H2* of T and D were comparable. The effect of category on H1*–H2* confirms previous findings on other languages with the four-way contrast [4, 5, 6]. However, the lack of effect of register on H1*–H2* is in opposition to findings from [21] which found vowels in Japanese IDS were significantly breathier than in ADS. Breathy voice might, thus, be a language-specific property of Japanese IDS rather than a general property of IDS. It is possible that the H1*–H2* at vowel midpoint is breathier than at vowel onset in Bengali IDS, but this possibility is not explored here and is left as an avenue for future research.

4. CONCLUSION

In summary, the acoustic manifestation of Bengali IDS revealed an asymmetry among the laryngeal categories in terms of VOT and PVI. In terms of primary acoustic cues to the four-way contrast, these results suggest that voiced-aspirated stops are cued by VOT and PVI since both of these were exaggerated in the hyper-articulated register. Similarly, voiced-unaspirated stops and affricates are cued by VOT. However, voiceless-aspirated stops and affricates, which are thought to be adequately cued by positive VOT alone, were not exaggerated in IDS. Finally, breathy voice was not exaggerated in IDS. This suggests that IDS may not be hyper-articulated across the board and that language-specific phonetic grammars may modulate the acoustic manifestation of different speech styles.

These results can also be viewed through the lens of phonological specification. Assuming that phonetic cues associated with specified features are exaggerated in hyper-articulated registers [22], these results support the representation of the Dh category with the features [voice] and [spread], the D category with [voice], and the T category as unspecified. However, since the VOT of the Th category was not exaggerated in IDS, this also supports the specification of this category with a feature other than [spread] (or none at all). Given that this result is problematic for most theories of laryngeal representation, the acoustic properties of registers like IDS or slow speech cannot always adequately comment directly on questions of phonological representation. The relationship between phonetic realization and phonological specification can be abstract, and evidence for featural representation must be *phonological*, not just phonetic.

The acoustic properties of IDS are language-dependent and hierarchically ordered. In some languages, vowels, but not consonants, are hyper-articulated [5]. This paper shows that there are asymmetries even within a single class of consonants – caregivers do not hyper-articulate equally across the board. This raises questions regarding the timeline of acquisition of this contrast and suggests that IDS may not always facilitate phonetic learning as it does not contain uniformly enhanced cues to contrasts in the target language.

5. REFERENCES

[1] L. Lisker and A. Abramson, “A cross-language study of voicing in initial stops: Acoustical measurements,” *Word*, vol. 20, pp. 384–422, 1964.

[2] G. N. Clements and R. Khatiwada, “Phonetic realization of contrastively aspirated affricates in Nepali,” in *Proc. 16th ICPhS*, 2007, pp. 629–632.

[3] M. K. Rami, J. Kalinowski, A. Stuart, and M. P. Rastatter, “Voice onset times and burst frequencies of four velar stop consonants in Gujarati,” *The Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3736–3738, 1999.

[4] I. Dutta, *Four-way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt*. University of Illinois at Urbana-Champaign, 2007.

[5] K. H. Berkson, “Capturing breathy voice: Durational measures of oral stops in Marathi,” *Kansas Working Papers in Linguistics*, 2012.

[6] J. Schertz and S. Khan, “Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu,” *J. Phon.*, vol. 81, 2020.

[7] P. K. Kuhl, J. E. Andruski, I. A. Chistovich, L. A. Chistovich, E. V. Kozhevnikova, V. L. Ryskina, E. I. Stolyarova, U. Sundberg, and F. Lacerda, “Cross-language analysis of phonetic units in language addressed to infants,” *Science*, vol. 277, no. 5326, pp. 684–686, 1997.

[8] U. Sundberg, “Consonant specification in infant-directed speech. Some preliminary results from a study of voice onset time in speech to one-year-olds,” *Working papers/Lund University, Department of Linguistics and Phonetics*, vol. 49, pp. 148–151, 2001.

[9] M. S. Fish, A. García-Sierra, N. Ramírez-Esparza, and P. K. Kuhl, “Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT,” *J. Phon.*, vol. 63, pp. 19–34, 2017.

[10] U. Sundberg and F. Lacerda, “Voice onset time in speech to infants and adults,” *Phonetica*, vol. 56, no. 3-4, pp. 186–199, 1999.

[11] K. T. Englund and D. M. Behne, “Infant directed speech in natural interaction—Norwegian vowel quantity and quality,” *Journal of psycholinguistic research*, vol. 34, no. 3, pp. 259–280, 2005.

[12] T. Benders, S. Pokharel, and K. Demuth, “Hypo-articulation of the four-way voicing contrast in Nepali infant-directed speech,” *Language Learning and Development*, vol. 15, no. 3, pp. 232–254, 2019.

[13] K. Yu, S. D. Khan, and M. Sundara, “Intonational phonology in Bengali and English Infant-directed speech,” *Proceedings of Speech Prosody 7*, pp. 1130–1134, 2014.

[14] S. D. Khan, “Bengali (Bangladeshi standard),” *J. IPA*, pp. 221–225, 2010.

[15] P. Boersma and D. Weenink, “Praat: Doing phonetics by computer [Computer program]. Version 6.3.03,” retrieved 17 December 2022 from http://www.praat.org/, 2022.

[16] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, “Voicesauce: A program for voice analysis,” in *Proc. 17th ICPhS*, 2011, pp. 1846–1849.

[17] D. Bates, M. Maechler, B. Bolker, and S. Walker, “lme4: Linear mixed-effects models using eigen and s4,” 2015, r package version 1.1-7. [Online]. Available: http://CRAN.R-project.org/package=lme4

[18] R Core Team, “R: A language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria, 2021.

[19] T. Cho and P. Ladefoged, “Variation and universals in VOT: evidence from 18 languages,” *J. Phon.*, vol. 27, no. 2, pp. 207–229, 1999.

[20] T. Cho and P. Keating, “Effects of initial position versus prominence in English,” *J. Phon.*, vol. 37, no. 4, pp. 466–485, 2009.

[21] K. Miyazawa, T. Shinya, A. Martin, H. Kikuchi, and R. Mazuka, “Vowels in infant-directed speech: More breathy and more variable, but not clearer,” *Cognition*, vol. 166, pp. 84–93, 2017.

[22] J. Beckman, P. Helgason, B. McMurray, and C. Ringen, “Rate effects on Swedish VOT: Evidence for phonological overspecification,” *J. Phon.*, vol. 39, no. 1, pp. 39–49, 2011.