MANNER DIFFERENCES IN THE PUNJABI DENTAL-RETROFLEX CONTRAST: AN ULTRASOUND STUDY OF TIME-SERIES DATA

Alexei Kochetov[1], Matthew Faytak[2], Kiranpreet Nara[1]

[1]University of Toronto, [2]University of California, Los Angeles
al.kochetov@utoronto.ca, faytak@g.ucla.edu, kiranpreet.nara@mail.utoronto.ca

ABSTRACT

Previous analysis of Punjabi ultrasound data suggests that the dental-retroflex contrast during consonant closure is less distinct in nasals than in stops. However, the contrast could potentially be encoded by tongue movement during the preceding vowel. Here, we analyse series of ultrasound frames extracted from the onset the preceding /a/ to the release of /t ʈ n ɳ/. Similarity of the contrast across manners was assessed by reducing the dimensionality of these series of frames and submitting them to a linear discriminant analysis, training on one manner and attempting to classify place in its counterpart. The results suggest that classifiers trained on stops are more accurate than those trained on nasals, and that classifiers trained on one manner do not generalize well to the other. This confirms that the dental-retroflex contrast in Punjabi nasals is less distinct than in stops, even taking the transition from the preceding vowel into account.

Keywords: speech production, articulation, retroflex, manner, ultrasound, Punjabi.

1. INTRODUCTION

Like most languages of South Asia [1], Punjabi (Indo-Aryan) exhibits a phonemic contrast between dental and retroflex place in consonants, cutting across various manners of articulation, including stops and nasals (e.g. baːt 'saying' vs. vaːʈ 'distance'; kʰaːn 'Khan (surname)' vs. kʰaːɳ 'mine') [9, 19]. The place contrast in Punjabi nasals, however, is more phonetically reduced, lesser in magnitude, and, consequently, is subject to variability and merger (in some dialects) [23]. The nasals' lesser distinctiveness is likely rooted in the physiological and aerodynamic factors involved in their production [17]. In general, coronal nasals have weaker constrictions than stops and show greater propensity to flapping (for /ɳ/) and apicality (for [n]) (see [5, 16] on the contrast in Gujarati and Kannada). These manner-specific constraints are at least in part counterbalanced by considerations of gestural economy and uniformity [16], as it is advantageous for speakers to employ the same lingual gestures for similar articulations, such as retroflexes or dentals as a class. Individual Punjabi speakers, however, may respond to these conflicting pressures in their own ways, producing phonetic variation across the population (cf. [13, 4, 8]).

The lesser distinctiveness and variability of the dental-retroflex contrast in nasals was clearly observed in [15]. This study analysed ultrasound data from Punjabi speakers producing the consonants /t, ʈ, n, ɳ/. Ultrasound video frames were extracted from the point of maximum tongue displacement and submitted to a principal component analysis (following [11, 12, 8]). The resulting scores were used to train a linear discriminant analysis (LDA) to classify new tokens by place of articulation (dental or retroflex). Two LDAs were performed, varying the training and testing sets by (1) training the model on stops and testing it on nasals, and (2) training on nasals and testing on stops. The results showed high accuracy in classification of the contrast in the training data, although higher for stops than nasals (100% vs. 92%). The classifier performed worse when extended to test data of the other manner: on average 67% of stops and 57% of nasals were classified correctly based on training sets of nasals or stops, respectively. These findings were interpreted as reflecting differences in the implementation of the place contrast across manners (cf. [6, 7] on Hindi), as well as the weaker contrast in nasals.

However, the difference in classification accuracy of stops and nasals could be due to temporal under-sampling of the data. A single frame from the consonant closure may not have captured the full realization of the contrast, particularly for inherently dynamic articulations such as retroflexes. Previous studies have shown that the tongue tip for retroflexes begins to retract early during the preceding vowel and, after achieving the target, continues to move forward during and after its closure [21, 17]. It is therefore plausible that some information about the dental-retroflex contrast in nasals is contained in the preceding vowel and was missed in the previous analysis. The goal of the current study is to extend the analysis of the Punjabi dataset by analysing the entire VC interval for /t, ʈ, n, ɳ/. The analysis also included two additional improvements in pre-processing of the data for analysis: using a more sophisticated noise-reduction filter and limiting the data submitted to PCA and LDA to a region of interest defined for each individual participant.

2. METHOD

2.1. Participants and materials

The participants and materials were the same as in [15]. That is, 14 native speakers of Punjabi (P3-P16, 7 females) from Punjab, India (residing in Toronto, Canada) read a list of nonsense words, including items with the 4 consonants in the symmetrical /a_aː/ context (adapted from [6, 7]), as shown in Table 1. The list was presented in the Gurmukhi script, which has distinct symbols for dental and retroflex stops and nasals. On average, 9 repetitions per item per speaker were collected with a total of 517 tokens.

Table 1: The words used in the study and numbers of analyzed word tokens and ultrasound frames.

Consonant	Word (Gurmukhi)	Word (IPA)	Tokens
dental stop	ਬਾਤਾਬ	[bataːb]	131
retroflex stop	ਬਾਟਾਬ	[baʈaːb]	130
dental nasal	ਬਾਨਾਬ	[banaːb]	127
retroflex nasal	ਬਾਣਾਬ	[baɳaːb]	129

2.2. Instrumentation and procedure

Data were collected using the *Telemed Echo Blaster* 128 CEXT-1Z system with an Articulate Instruments pulse-stretch unit. The frame rate was 35.39 fps., the probe field of view and depth were 93.27 degrees and 150 mm. The probe was stabilized using an Articulate Instruments stabilization headset [24]. The audio, collected at a sampling rate of 22,050 Hz, was synchronized with the ultrasound video using the *Articulate Assistant Advanced* software [2]. The sessions were conducted in Punjabi.

2.3. Analysis

For each token, frames were extracted for the entire VCV interval, that is, the target consonant closure and the preceding and following vowels. VC transitions are known to be important for encoding the dental-retroflex contrast [5, 21]. Given this, we selected for the analysis four evenly spaced frames from the first half of each time series. This subset typically included the closure and release, as speakers produced the first vowel shorter than the second vowel (e.g. [baʈaːb]), following Punjabi phonotactics [19].

The extracted frames were filtered to reduce speckle noise following [20]. All frames were cropped to a region of interest, manually drawn for each speaker, that contained all tongue contours. Each token's four cropped frames were then concatenated into a single basis object (Figure 1).

Following [11, 20], a principle component analysis (PCA) was carried out on each speaker's set of series objects to reduce the dimensionality of the data. Rather than carry out a single PCA including all speakers, a separate PCA was carried out for each speaker to avoid incorporating non-linguistic speaker variation (due to probe orientation and morphological variation) into the speaker-specific realization of the linguistic contrasts being examined. Loading plots of PCs 1-3 or "eigentongues" [11] for one speaker are shown in Figure 2.

Figure 1: A sample image of four selected frames for [banaːb] by P14.

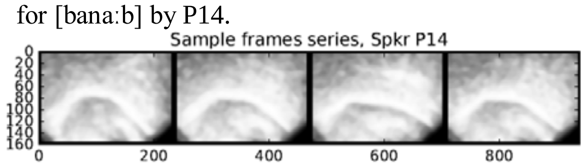
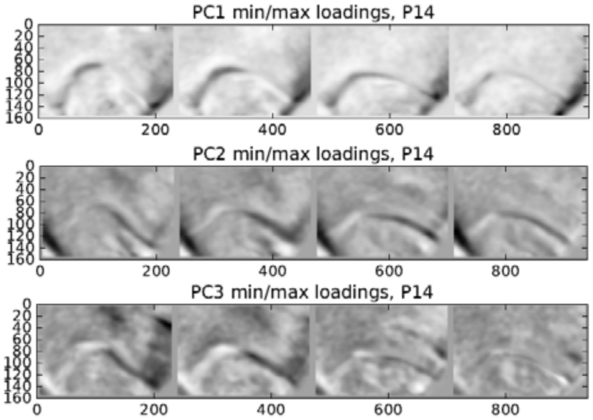


Figure 2: PC1-3 "eigentongues" (explaining over 50% of variation) of four evenly-spaced frames for Speaker P14.



Taking the PC scores as a representation of tongue shape, two linear discriminant analyses (LDAs) were performed for each speaker as follows. A first LDA was trained using PC scores for the dental and retroflex stops /t, ʈ/; the resulting linear discriminant was used to classify the dental and retroflex nasals /n, ɳ/ as either dental (/t/) or retroflex (/ʈ/) in terms of tongue shape. A second LDA reversed the roles of each manner, training the model on the nasals /n, ɳ/ and classifying the stops /t, ʈ/ as either /n/ or /ɳ/. The two linear discriminants served as indices of the speaker-specific dental-retroflex contrast for the manner each was trained on. Classifying one manner according to the linear discriminant trained on the other in turn indicated how comparable the dental-retroflex contrast was across manners for a given speaker. The LDA classification results are presented below as rates of 'dental' and 'retroflex' responses. PCAs and LDAs were carried out using scikit-learn [22], and "eigentongue" images were produced using Matplotlib [12], both done in Python.

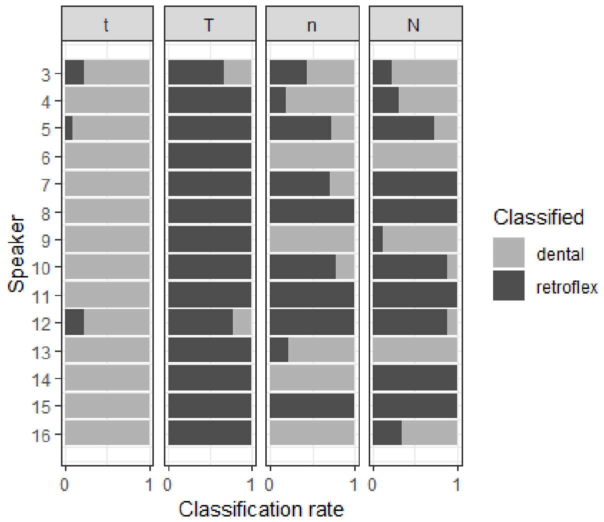
3. RESULTS

3.1. Training based on stops

Figure 3 shows results of the first LDA, which was trained on stops in order to classify nasals. Classifications of training (/t, ʈ/) and test categories (/n, ɳ/) are shown as proportions of 'dental' and 'retroflex' responses, separately for each speaker's dataset. It can be seen that the training categories /t/ and /ʈ/ produced by all speakers were classified overwhelmingly correctly. Only speakers P3, P5, and P12 showed less than perfect classification of the stops. Some of these errors can be attributed to image sets where the tongue front was partially obscured by the chin bone (particularly for P12).

The testing categories, on the other hand, were classified differently for different speakers. Only six speakers showed the expected /n/ classification – as predominantly dental (P4, 6, 9, 13, 14, 16). Other seven speakers showed incorrect, predominantly 'retroflex' responses (P5, 7, 8, 10, 11, 12, 15). Speaker P3 showed highly variable classifications. The classification of /ɳ/ was also variable, while showing higher rates for the correct category: it was identified as predominantly retroflex for seven speakers and as predominantly dental for six speakers. Curiously, classifications of testing categories (as dental or retroflex) for most speakers were similar regardless of the consonant. For example, both /n/ and /ɳ/ by P6 were uniformly classified as dental, while both /n/ and /ɳ/ by P8 were uniformly classified as retroflex. As a result, the place contrast in nasals has failed to be distinguished by LDA, except for a single speaker (P14).

Figure 3: LDA classification of 4 consonants: trained on stops, testing nasals; 'T', 'N' = /ʈ/, /ɳ/.

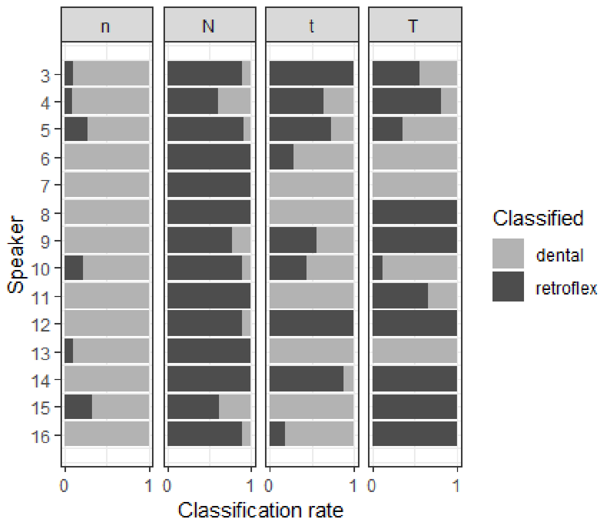


The results of place classification for nasals can be interpreted in two ways. First, speakers (other than P14) neutralize the place contrast to either /n/ or /ɳ/. The overall higher number of retroflex classifications suggests that neutralization is towards the retroflex rather than the dental place. Or, it could be an intermediate category that is identified by LDA as more similar to /ʈ/ than /t/. Another possibility is that the speakers did produce the contrast, yet it was realized differently in nasals than in stops.

3.2. Training based on nasals

Figure 4 shows results of the second LDA, which was trained on nasals in order to classify stops. The classification of the training categories, /n, ɳ/, was relatively accurate, but lower than for stops in Figure 3. It was nevertheless well above chance and observed for all speakers. This is an important finding, as it indicates that place in nasals is produced largely distinctly, contrary to one of the possible conclusions reached above in section 3.1. The classification of the testing categories /t/ and /ʈ/ was much poorer and showed considerable variation. Thus, /t/ and /ʈ/ were classified predominantly correctly for seven and eight speakers, respectively, while they were misclassified or variably classified for the other speakers. Despite this variability, the results here show a better discrimination of the place contrast in nasals: the model was able to correctly classify test segments' place for four speakers – with a 100% accuracy for P8 and P15, and at relatively high accuracy rates for P11 and P16). This is different from the results of the first LDA, which was able to distinguish the contrast for only one speaker.

Figure 4: LDA classification of 4 consonants: trained on nasals, testing stops; 'T', 'N' = /ʈ/, /ɳ/.

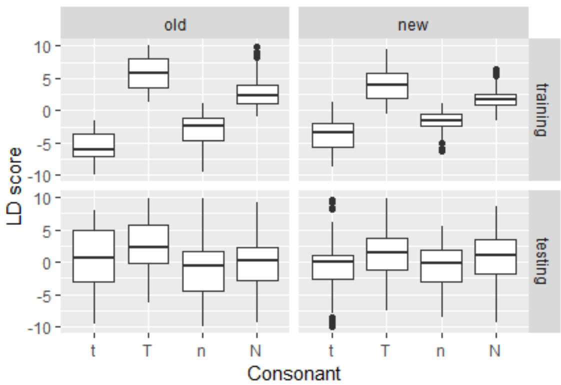


Overall, the results for this set support our second interpretation made based on the stop training set: Punjabi speakers do not seem to neutralize the place contrast in nasals; rather they appear to implement the contrast differently in nasals than in stops.

3.3. Comparison with previous analysis

The conclusions we reached in the previous sections are that (i) the dental-retroflex contrast in Punjabi is implemented somewhat differently in stops and nasals and (ii) the contrast in nasals is not as robust as in stops. Exactly the same conclusions were reached in [15], the study that analysed the same data set by examining a single ultrasound frame per token. In contrast, here we analysed a set of four frames from the entire VC interval of each token. A question arises as to whether there are any differences in the results between the two analyses. Figure 5 presents a comparison of linear discriminant scores of the two analyses, separately by consonant and task. Dentals and retroflexes are expected to have lower and higher LD scores, respectively. Of interest for us here is the degree of separation between the consonant categories under each task condition and in each analysis. The figure confirms that the two analyses were very similar in distinguishing the contrast in the training sets (although better for stops than nasals) and largely neutralizing it in the testing set (while doing slightly better in stops). The new analysis, however, shows somewhat lower discrimination of training categories and higher variability in test categories compared to the old one.

Figure 5: LD scores for 4 consonants comparing two analyses (‘old’ and ‘new’); ‘T’, ‘N’ = /ʈ/, /ɳ/.



These differences can be attributed to differences in frame selection. Frames for the current analysis were selected automatically (at four equidistant points within the VC interval), and thus may not have included the frames where the tongue was maximally displaced. In contrast, such frames were manually selected in the old analysis, and appear to have contributed crucially to the discrimination of the place contrast. In the future, it would be useful to combine these two methods or possibly increase the number of analysed frames.

4. DISCUSSION

The goal of this study was to extend the analysis of the Punjabi dental-retroflex contrast explored in [15] by including ultrasound frames from the VC interval for the consonants /t, ʈ, n, ɳ/. This was done because VC transitions have been noted to play an important role in distinguishing the dental-retroflex contrast [5, 21]. One may therefore expect that the lack of cross-manner generalizations observed in [15] could have resulted from the absence of this information. The results of the current study, however, are remarkably similar to those in [15] and point to considerable differences in the phonetic implementation of the place contrast in stops and nasals. They also confirm the earlier observation that the contrast in nasals is not as distinct as in stops and highly variable (see also [19] on Punjabi; [5, 7] on Gujarati and Hindi). Importantly, however, results of both studies lead to a conclusion that, despite this variation, the place contrast in nasals is not (fully) neutralized by any of our speakers. This is contrary to the authors’ auditory impressions reported in [15], where some speakers were perceived to merge the two nasals or to vary them across repetitions. If this were indeed the case, we would have observed considerably lower correct classification rates for nasals in the training set. It appears thus that the contrast in nasals is maintained by all speakers, but to various degrees and not always consistently [cf. 13, 4]. At the same time, we should not discount a possibility of category separation being somewhat influenced by spurious differences in images (e.g. probe rotation across repetition blocks).

Regardless of the exact interpretation of the results and some differences between the methods employed, the data clearly exhibit substantial variation within and across speakers, and this variation is primarily observed in nasals rather than stops. This variation can be interpreted as manifestation of individual strategies to resolve conflicting demands – accommodating manner-specific physiological and aerodynamic constraints [17] and enforcing gestural uniformity [18, 16, 3, 8]. A few speakers appear to prioritize uniform implementation of the retroflex-dental contrast, but it is obvious that this is rather exceptional among our speakers, suggesting that the change towards the reduction of the contrast in the language is fairly advanced.

ACKNOWLEDGEMENTS

Thanks to Maida Percival and Punjabi speakers. This research was supported by a *Social Sciences and Humanities Research Council of Canada* grant.

5. REFERENCES

[1] Arsenault, A. 2017. Retroflexion in South Asia: Typological, genetic and areal patterns. *J. of South Asian Languages and Linguistics,* 4, 1–53.

[2] Articulate Instruments Ltd., Articulate Assistant Advanced User Guide: Version 2.16. Edinburgh, UK: Articulate Instruments Ltd., 2012.

[3] Chodroff, E., Wilson, C. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *J. of Phonetics,* 61, 30–47.

[4] Clopper, C. G. 2014. Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Laboratory Phonology*, 5, 69–90.

[5] Dave, R. 1977. Retroflex and dental consonants in Gujarati: A palatographic and acoustic study. *Annual Report of the Institute of Phonetics, University of Copenhagen (ARIPUC),* vol. 11, 27–156.

[6] Dixit, R. P. 1990. Linguotectal contact patterns in the dental and retroflex stops of Hindi. *J. of Phonetics*, 18, 189–201.

[7] Dixit, R. P. 1999. Tongue-palate contact during retroflex consonants /D/, /R/, /N/ in Hindi. In: Braun, A. (ed.), *Advances in Phonetics: Proc. of the International Phonetic Sciences Conference (IPS),* Bellingham, WA, June 27-30, 1998, vol. 106, Franz Steiner Verlag.

[8] Faytak, M. D. 2018. Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese. UC Berkeley. Retrieved from https://escholarship.org/uc/item/47j8969j

[9] Gill, H. S., Gleason, H. A. 1969. *A Reference Grammar of Punjabi.* Department of Linguistics, Punjabi University.

[10] Hoole, P., Pouplier, M. 2017. Öhman returns: New horizons in the collection and analysis of imaging data in speech production research. *Computer Speech and Language*, 45, 253–277.

[11] Hueber, T., Aversano, G., Cholle, G., Denby, B., Dreyfus, G., Oussar, P., Roussel, P., Stone, M. 2017. Eigentongue feature extraction for an ultrasound-based silent speech interface. *Proc. of ICASSP '07*, Honolulu, HI, I-1245–I-1248.

[12] Hunter, J. 2007. Matplotlib: A 2D graphics environment, *Computing in Science & Engineering*, 9, 90–95.

[13] Johnson, K., Ladefoged, P., Lindau, M. 1993. Individual differences in vowel production. *J. of Acoustical Society of America*, 94, 701–714.

[14] Keating, P. 2003. Phonetic and other influences on voicing contrasts. *Proc. of the 15th ICPhS*, 375–378.

[15] Kochetov, A., Faytak, M., Nara, K. 2018. The retroflex-dental contrast in Punjabi stops and nasals: A principal component analysis of ultrasound images. *Proc. Interspeech 2018*. 5 pp. https://www.isca-speech.org/archive/Interspeech_2018/pdfs/1457.pdf

[16] Kochetov, A., Sreedevi, N., Kasim, M., Manjula, R. 2014. Spatial and dynamic aspects of retroflex production: An ultrasound and EMA study of Kannada geminate stops. *J. of Phonetics,* 46, 168–184.

[17] Kochetov, A., Tabain, M., Sreedevi, N., Beare, R. 2018. Manner and place differences in Kannada dentals and retroflexes: Articulatory and acoustic results. *J. of Acoustical Society of America,* 144, 3221–3235.

[18] Maddieson, I. 1996. Gestural economy. *UCLA Working Papers in Phonetics,* 94, 1–6.

[19] Malik, A. N. 1995. *The Phonology and Morphology of Panjabi.* New Delhi: Munshiram Manoharlal Publishers.

[20] Mielke, J., Carignan, C., Thomas, E. R. 2017. The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study. *J. of Acoustical Society of America*, 142, 332–349.

[21] Narayanan, S., Byrd, D., Kaun, A. 1999. Geometry, kinematics, and acoustics of Tamil liquid consonants. *J. of Acoustical Society of America*, 106, 1993–2007.

[22] Pedregosa, F., Varoquaux, G., Gramfort., A., Michel, V., Thirion, B., … Duchesnay, E. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

[23] Sandhu, B. S. 1986. *The Articulatory and Acoustic Structure of the Panjabi Consonants.* Patiala: Publication Bureau, Punjabi University.

[24] Scobbie, J. M., Wrench, A., van der Linden, M. 2008. Head-probe stabilization in ultrasound tongue imaging using a headset to permit natural head movement," *Proc. of the 8th International Seminar on Speech Production*, 373–376.