A preliminary ultrasound study of Nepali lingual articulations

Alexei Kochetov, University of Toronto
Marianne Pouplier
Sarah Truong, University of Toronto

Available at: https://works.bepress.com/alexei_kochetov/35/
ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)

5aSCb57. A preliminary ultrasound study of Nepali lingual articulations
Alexei Kochetov*, Marianne Pouplier and Sarah Truong

*Corresponding author’s address: The Department of Linguistics, The University of Toronto, Toronto, M5S 3G3, Ontario, Canada, al.kochetov@utoronto.ca

Previous descriptive and phonetic works on Nepali provided conflicting accounts of place contrasts in coronal consonants. Specifically, apical stops were characterized as either retroflex or alveolar, while laminal affricates were described as either alveolar or palatal. Some of these works used static palatography, which shows the contact between the tongue and the palate, but provides no information about the tongue shape for a given consonant or its dynamic properties. In this study we used ultrasound to image tongue shapes for various Nepali lingual consonants produced by a single native speaker of Brahmin dialect. The results showed that the speaker’s apical stops were produced with a substantially raised tongue front and retracted tongue tip, as would be expected of retroflex articulations. Laminal affricates had the tongue shape similar to dental stops, yet with a somewhat retracted tongue tip, indicative of the alveolar constriction. Apicals that differed in laryngeal features (voiceless, voiced, aspirated, breathy) did not show systematic differences in the tongue shape, except for the voiced stop, which was somewhat less retracted. While limited to the single speaker, the results confirm and extend some previous observations about Nepali coronals as showing a 3-way place contrast among dentals, alveolars, and retroflexes.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Nepali, an Indo-Aryan language spoken mainly in Nepal by over 11 million people (Lewis, 2009), has a complex set of coronal obstruents. Among these are stops and affricates that have traditionally been described as ‘dental’, ‘retroflex’, and ‘palatal’, and consonants of each of these places can be either voiceless, voiceless aspirated, voiced, or breathy voiced. These place and laryngeal contrasts are illustrated in Table 1 (adapted from Acharya, 1991: 21-22; cf. Khatiwada, 2009).

<table>
<thead>
<tr>
<th>Place</th>
<th>laminal ‘dental’</th>
<th>apical ‘retroflex’</th>
<th>laminal ‘palatal’</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiceless</td>
<td>tal ‘lake’</td>
<td>tţok ‘bite (imp.)’</td>
<td>tţala ‘movement’</td>
</tr>
<tr>
<td>voiceless aspirated</td>
<td>tţal ‘plate’</td>
<td>tţok ‘hit (imp.)’</td>
<td>tţala ‘leather’</td>
</tr>
<tr>
<td>voiced</td>
<td>dan ‘donation’</td>
<td>dţoka ‘bamboo baskets’</td>
<td>dzari ‘alimony’</td>
</tr>
<tr>
<td>breathy voiced</td>
<td>dţan ‘rice (unhusked)’</td>
<td>dţoka ‘door’</td>
<td>dţari ‘pitcher’</td>
</tr>
</tbody>
</table>

Previous descriptive and phonetic works on Nepali provided conflicting accounts of the exact place and manner of the retroflex and palatal consonants. For example, Acharya (1991: 16) noted that Nepali ‘retroflexes’ are more fronted and do not involve as much curling of the tongue tip as retroflexes in other Indo-Aryan languages (such as Hindi); he preferred to refer to these consonants as ‘alveopalatal’. Pokharel (1989: 83) also remarked that Nepali retroflexes “are not generally retroflexed”, but are rather ‘apico-alveolar’, similar to the English /t/ and /d/. His palatograms for the words with geminates /tţ/ (/mu/tţ/ and /pa/tţa/; p. 298 ff.), however, showed a considerably retracted, post-alveolar constriction, very different from the more front laminal denti-alveolar /t/. More recently, Khatiwada (2007) obtained static palatograms and linguograms of Nepali retroflexes /tţ dţ dţţ/ from 7 speakers. The results showed that, at least in the back vowel context, retroflexes were consistently post-alveolar, while showing some variation in the active articulator – the tip (apical) or the underside of the tongue (sub-apical). He also found that sub-apical realizations were somewhat more common in voiced and breathy voiced retroflexes, compared to the voiceless and voiceless aspirated ones. Overall, this shows that Nepali retroflexes can involve the curling of the tongue similar to retroflexes in other languages of South Asia (see e.g. Ladefoged & Bhaskararao, 1983), contrary to previous generalizations.

With respect to the ‘palatals’, these consonants have been described by some researchers as stops produced with the tongue blade against the front part of the hard palate and involving ‘the minimal amount’ of affrication (i.e. laminal alveolo-palatal stops /c cţ j j/; Acharya, 1991: 16-18). In contrast, Pokharel (1989: 82-84) described these consonants as affricates articulated with the blade of the tongue against the alveolar ridge (laminal alveolars), and this was confirmed by his palatographic study. Similar conclusions about these sounds were drawn by Clements & Khatiwada (2007) from palatograms and linguograms obtained from four speakers of the language. Given this, the ‘palatals’ appear to be more appropriately described as the laminal alveolar /s sţ dţ dţţ/ (Khatiwada, 2009).

The goal of this study is to explore the use of a different articulatory method to investigate place differences in Nepali coronals. We use ultrasound to image the tongue shape during the production of Nepali dental, retroflex, and ‘palatal’ stops/affricates that differ in their laryngeal features. Ultrasound data of this kind can provide novel spatial and dynamic information about these consonant articulations, complementing the available static palatographic data on the location and type of the tongue-palate constriction.

METHOD

Participant and Materials

A single speaker of Brahmin dialect of Nepali from Kathmandu participated in the experiment. He was in his early twenties and at the time of the experiment was a student at Yale University.

The stimuli included meaningless combinations of real Nepali words of the type ‘Pa CaP’ or ‘PaC aP’, where P is a labial /p/, /b/, or /m/, and C is a dental stop /t/, a ‘palatal’ affricate /ts/, a velar stop /k/, or a retroflex stop of any of the four laryngeal specifications, /tţ dţ dţţ/. These stimuli, shown in Table 2, formed two sets: Set 1 was designed to investigate place differences in various voiceless stops and affricates, as shown in; Set 2 was designed to investigate the effect of laryngeal features on the realization of the retroflex place, as shown in Table 3. Both sets
included target consonants in syllable-initial and syllable-final position, making it possible to examine the role of the syllable in the realization of these consonants. While retroflexes and ‘palatals’ were the main focus of the study, dental and velar stops were included to provide a more anterior and more posterior reference points for the other two places. The words for the stimuli were selected from Acharya’s (1991) grammar of Nepali and online dictionary sources. The full set of stimuli examined in this study is presented in Table 2. All the stimuli were placed in a carrier phrase /ma _ p`eri b`a`n`u/ ‘I say ___ again’, randomized, and presented in the Devanagari script on a computer screen.

<table>
<thead>
<tr>
<th>Onset</th>
<th>Coda</th>
</tr>
</thead>
<tbody>
<tr>
<td>dental</td>
<td>retroflex</td>
</tr>
<tr>
<td>vls.</td>
<td>ba tap</td>
</tr>
<tr>
<td>vls. asp.</td>
<td>ba ċap</td>
</tr>
<tr>
<td>vd.</td>
<td>ba ċam</td>
</tr>
<tr>
<td>breathy vd.</td>
<td>ba ċap</td>
</tr>
</tbody>
</table>

**Instrumentation and Procedure**

The data were collected at Haskins Laboratories, New Haven, Connecticut, using an Aloka SSD-1000 ultrasound system with a 3-5 MHz probe and a Sennheiser shotgun microphone. The participant was seated in a specialized dental chair with a head rest, while holding the probe under his chin to capture a sagittal view of the tongue. The stimuli were presented on a computer screen, with the list being repeated 5 times. This was expected to result in 90 tokens in total (14 stimuli * 5 repetitions). However, one token of /ba tap/ was omitted by the speaker; 2 other tokens (of /ba ċam/) involved false starts or hesitations – resulting in 87 tokens available for the analysis.

**Analysis**

Frames from the ultrasound video were also extracted using MPEG Streamclip freeware (version 2.1; Cinque, 2011) at 30 frames per second. Two frames for each item were selected for the analysis: the frame of the maximum displacement of the tongue in the expected direction (e.g. the frontmost for the dental and the highest central for the velar) during the acoustic closure of the consonant, and the frame occurring 5 frames (150 ms) away from the point of maximum displacement. These will be referred to as the ‘max frame’ and ‘-5 frame’. The latter corresponded to the tongue configuration during the low central/back vowel /a/. This point was assumed to be very similar across all the items, and was used to determine the direction of the movement for each consonant. Figure 1 shows a set of sample frames for the item /ba ċap/.

**FIGURE 1.** Frames extracted from ultrasound video for a sample token of /ba ċap/. The tongue tip is on the right.

The freeware EdgeTrak (version 1.0.0.2; Li, Kambhamettu, & Stone, 2005) was used to semi-automatically trace tongue contours for each frame. Traced contours were extracted as 100 X and Y coordinates (with pixels converted to millimeters) and saved in a tab-delimited format. Coordinates were further imported into R (version 2.15.1; Team, 2012) for analysis using a series of Smoothing Spline Analyses of Variances (SS-ANOVAs). This method determines whether two curves are significantly different from each other at any particular interval. Visually significant differences are indicated by the lack of overlap of 95% confidence intervals of the two data sets (Gu, 2002; Davidson, 2006). Separate SS-ANOVAs were run for the max frames of set 1 items (/t/ vs. /k/, /t/ vs. /s/, /l/ vs. /k/, /s/ vs. /l/, /s/ vs. /k/) and Set 2 items (/t/ vs. /t/, /l/ vs. /l/, /l/ vs. /l/). In addition, SS-ANOVAs were run comparing the max and -5 frames for the same consonant (e.g. max frame for /t/ and -5 frame for /t/).
RESULTS

Overall patterns

Figure 2 presents mean contours for various items (averaged over 5 tokens). The plot in (a) shows contours 5 frames before the consonant constriction. All four contours are very similar, yet the visible part of the tongue tip/blade for the dental extends much further than for the other consonants. This could be either due to the anticipatory fronting of the tongue tip for /t/ or its retraction for /q/, /h/, and /k/. The plot in (b) shows the tongue contours for the same four consonants at the maximum constriction. The tongue for the dental is in a front and moderately raised position, presumably articulating against the upper teeth and the anterior part of the alveolar ridge. The front part of the tongue is considerably raised for the retroflex, with the tip likely touching behind the alveolar ridge. Note that ultrasound does not generally capture the tongue tip proper, particularly when its surface is near-parallel to ultrasound beam (Stone, 1995). Interestingly, the tongue shape for the ‘palatal’ closure is almost identical to the dental, yet with the tip/blade being in a more retracted position, presumably articulating against the alveolar ridge. The velar constriction, as expected, involves a substantial raising of the tongue dorsum, which results in the retraction of the tongue tip/blade. The plot in (c) shows four retroflex consonants that differ in laryngeal features, taken at the point of the maximum constriction. All four tongue shapes are similar, having the front of the tongue in a high and somewhat posterior position. The tongue shape for the voiced /q/, however, is somewhat more front and lower than the others, while the tongue dorsum/root for the voiceless aspirated /q/ is slightly more retracted. Finally, plot in (d) shows dynamic aspects of the dental and retroflex articulations, comparing their -5 and max frames. It is obvious that both dental and retroflex involve a considerable fronting of the entire tongue relative to the position for /a/. They also show raising of the tongue tip/blade, which, for the retroflex consonant, is much greater, and accompanied by retraction.

The following sections examine whether the observed differences among the consonants and frames are statistically significant, and in which particular regions of the tongue. This is done separately first for place contrasts, for laryngeal contrasts, for different time-points, and for onset and coda positions.
Place contrasts

Figure 3 shows results of SS-ANOVAs comparing the voiceless retroflex stop constriction to its dental, velar, and palatal counterparts, based on onset and coda tokens pooled together (i.e. 10 tokens per consonant). Dotted lines delineate the areas where the two contours show no overlap of their 95% confidence intervals, and therefore significantly different. Thus, /t/ and /t/ are not significantly different in terms of the back/root of the tongue, but are significantly different throughout the rest of the tongue shapes. The same can be observed for the retroflex and palatal contours. On the other hand, the tongue back/root for the retroflex is significantly more front (and lower) compared to the velar stop.

![Graphs showing Place contrasts](image)

**FIGURE 3.** SS-ANOVA comparisons of the tongue shapes during the consonant constriction (max frames) of /t/ (‘T’) vs. /t/, /k/ (velar), and /s/ (‘c’). The tongue tip is on the right. The retroflex curve is indicated in black; the scale is in mm.

Laryngeal contrasts

Comparisons among pairs of retroflexes with different laryngeal settings (onset and coda tokens combined), shown in Figure 5, reveal small significant differences, mainly in the degree of fronting of the tongue body, and, for /l/ vs. /l/ in raising of the tip/blade. It is not clear whether these differences represent actual consonant-specific articulatory differences or an artifact of a small sample size of the study. Note that there were only 3 tokens for the onset /l/, compared to 5 for the other consonants.
FIGURE 4. SS-ANOVA comparisons of the tongue shapes during the consonant constriction (max frames) of /s/ (‘c’) vs. /t/ and /k/. The retroflex curve is indicated in black; the scale is in mm.

FIGURE 5. SS-ANOVA comparisons of the tongue shapes during the consonant constriction (max frames) of /t/ (‘T’) vs. /th/ and /d/ (‘D’). The voiceless curve is indicated in black; the scale is in mm.
Temporal differences

Figure 6 presents results of SS_ANOVAs comparing two different time frames for /t/ and two different frames for /t/ (onset and coda tokens combined). The differences between the pairs of contours are significant throughout almost the entire contours, confirming the earlier observation that the tongue moves forward for both retroflexes and dentals.

Syllable position

Finally, the results of SS_ANOVAs performed to compare the same consonant in onset and coda (5 tokens each) are shown in Figure 7. It can be seen that there are minor, yet significant differences in the degree of fronting of the tongue (more front for the coda /t/ and /ts/ than for the same consonant in onset), backing of the tongue (more back for the coda /k/ than for the onset /k/), or raising for the tongue body (higher for the coda /t/ than for the onset /t/). Again, while these differences may reflect some genuine syllable position effects (see e.g. Krakow, 1999), they may as well be spurious effects of the small sample size (which is in fact reduced by half compared to the analyses presented above).

DISCUSSION AND CONCLUSION

The goal of this study was to examine tongue shape differences among Nepali coronal obstruents using ultrasound. As previous descriptive phonetic and static palatographic studies disagreed on the exact location of the consonants traditionally referred to as ‘retroflex’ and ‘palatal’, it was of particular interest to examine how these were different from each other and other lingual consonants. The results obtained from a single speaker of the Brahmin dialect of Nepali showed a robust contrast between dental and retroflex stops, at least in the low vowel context used in the study. While the tongue body was fronted for both, the tongue tip and the blade were substantially raised and retracted for the retroflex. While the ultrasound method did not permit us observe the curling of the tongue tip and its presumed contact behind the alveolar ridge, the overall tongue shape for our speaker’s retroflex stop was quite similar to the more proto-typical sub-laminal retroflexes of Kannada (Dravidian), as investigated in Kochetov, Sreedevi, Kasim, & Manjula (to appear) (cf. Ladefoged & Bhaskararao, 1983 on Telugu, Hindi, and Tamil, retroflex stops). Specifically, the Kannada speakers in this study had a similar fronting of the tongue body for the retroflex relative to the velar, and the raising and retraction of the tongue tip/blade relative to the dental. Overall, the current results, are therefore consistent with Khatiwada’s (2007) static palatography findings that Nepali retroflexes can involve some curling of the tongue tip, being apical or sub-apical post-alveolars (in the back vowel context). This contradicts the descriptions of retroflexes as more fronted apicals (‘alveopalatal’: Acharya, 1991; ‘alveolar’ Pokharel, 1989, but see his palatographic evidence to the contrary).
Our investigation of retroflex stops with different laryngeal settings revealed some relatively small differences. Specifically, we found that voiced /d̪/ had a somewhat more front tongue body and a lower tongue blade/tip than its voiced counterpart. This is different from Khatiwada (2009) who found a tendency towards a greater retroflexion for voiced stops than for voiceless ones (see Hamann & Fuchs, 2010 on the phonetic motivation of such differences). As we mentioned, one plausible explanation for our difference is the variability in the data and a particularly small size for this consonant (see above). It should be noted, however, that Nepali /d̪/ is known to vary freely with a retroflex flap [ɾ] in the intervocalic position (Acharya, 1991; Khatiwada, 2009), and therefore our tokens may have included some flaps which were either more anterior or were produced too fast for the ultrasound frame rate to capture.

Our results also revealed that the speaker’s realization of the ‘palatal’ affricate was very fronted, laminal, and likely alveolar (cf. Clements & Khatiwada, 2007). Its overall tongue shape was remarkably similar to the laminal dental, yet with the tip.blade being in a somewhat posterior position. This is very different from the Kannada alveolo-palatal affricate produced with the tongue front in a much higher position than for the dental stop (in Kochetov et al., to appear). Our informal examination of spectrograms of the ‘palatals’ also confirmed that these were affricates, with release portions as long as the half of the consonant duration. This agrees with Pokharel’s (1989) and Clements & Khatiwada’s (2007) characterization of these consonants, but contradicts Acharya’s (1991) observations. Most consonants showed some, albeit small, differences in syllable onset and coda positions. Whether these differences reflect more basic syllable position effects, requires a further, more focused investigation.

FIGURE 7. SS-ANOVA comparisons of the tongue shapes during the consonant constriction (max frames) of /ba ʔap/ (voiceless) vs. /ba ʰap/ (voiceless aspirated), /ba ɾam/ (voiced), and /ba ɾam/ (breathy voiced). The voiceless curve is indicated in black; the scale is in mm.
To conclude, while limited to the single speaker, the results confirm and extend some previous observations about Nepali coronal obstruents as including a laryngeally complex set of laminal dental stops, laminal alveolar affricates, and apical or sub-laminal retroflex stops.

ACKNOWLEDGMENTS

The authors would like to thank Darya Kavitskaya for help and advice at the earlier stage of the project. The work was supported by NIH Grant DC-02717 to Doug Whalen and Haskins Laboratories.

REFERENCES


