Production and perception of lexical tones in Deori

Prarthana Acharyya[1], *Shakuntala Mahanta*[1]

[1]Department of Humanities and Social Sciences, Indian Institute of Technology Guwahati, India

a.prarthana@iitg.ernet.in, smahanta@iitg.ernet.in

Abstract

Deori, a Tibeto-Burman language (henceforth TB) belonging to the Bodo- Koch language group [10] can be unambiguously called an endangered language [2]. This paper analyzes the production and perception of lexical tones in Deori. Most recently, some remnant lexical tonal contrast has been shown in Deori with acoustic evidence [11]. Following the minimal tonal contrast prevalent among the older generation [11] a production experiment was conducted to determine the tonal distinction in the speech of the younger generation. The production experiment was followed by a perception test to investigate the impact of language experience on the perception of lexical tones in Deori. The result of production test shows a trend of underlying tone reversal H>L; L>H in monosyllables and a considerable F0 variation in the disyllabic stems. The tone reversal can be described as tone swap where a high tone has a low fundamental frequency and vice-versa. The perception test result too shows a similar trend of tone reversal in monosyllables as the listeners tend to misidentify tones mainly the younger generation speakers. Our result indicates that social factors such as language experience play a role in the categorization of tones in Deori.

Index Terms: tone, production, perception, tone swap.

1. Introduction

Deori, a Tibeto-Burman language of the Bodo-Koch sub family is spoken in the eastern parts of Assam and some parts of Arunachal Pradesh. It has a population of 41,161 (2001 census). It is spoken by approximately 28,000 people [2]. Deori has been referred to as a “lesser known language [10] and appears to be rather deviant” and a “discrete group” [4]. Deori has been explicitly listed as a ‘definitely endangered’ language [14] and as a ‘severely endangered language’ [2].

It has been claimed that languages belonging to the Tibeto-Burman family range from having many tonal contrasts to none, and from displaying emerging tonal contrasts to disappearing ones [6]. While tonogenesis is a widely acclaimed feature in the literature on tone, a large body of literature also records a gradual loss of the already existing tone, a process known as tonoexodus [6]. In the existing literature on Deori there have been many observations about tone. Lexical items in Deori are differentiated either by tone or nasalization (apart from other segmentally different minimal pairs) [16]. It has been postulated that Deori exhibits three tonal distinctions which changes the lexical meaning of the word [13]. It has also been found that Deori has two tonal distinctions, but synchronically, loss of tonal distinctions is apparent in the language [4]. In a previous study [11] it has been shown that Deori manifest level tones, and the distinction is preserved in a limited set of words. It was stated that the domain of Tone Bearing Unit (henceforth TBU) in Deori is the entire phonological word rather than the syllable. In the current study we would like to investigate tone production and tone perception in Deori and investigate the initial stages of tonoexodus in progress in Deori mainly in the speech of the younger generation speakers.

2. Production Test

In order to investigate the current use of lexical tone distinctions in Deori and language change attested in the speech of the younger generation speakers a production experiment was conducted.

2.1. Materials

For the present study we used the same set of words that were examined in the previous study on tone [11] as the tonal contrasts for these words were already established in the previous study. As vowel nasalization is also a factor in distinguishing homophonous words in Deori, this experiment ruled out words with nasal-oral contrast. The monosyllabic words were of CV syllable type and the bisyllabic words were of CVCV and VCV syllable types as shown in Table 1. For each of the monosyllabic words there were total of 150 samples (5 words*2 tones*5 speakers*3 iterations) and for bisyllabic words there were 240 samples (8 words*2 tones*5 speakers* 3 iterations).

Table 1: *The dataset of homophonous words*

Deori	Gloss	Deori	Gloss
akɯ̃	‘ear’	akɯ̃	‘upland’
tʃĩjã	‘fish’	tʃĩjã	‘wife of younger brother’
uzɯ̃	‘navel’	uzɯ̃	‘bamboo tube’
tiri	‘banana’	tiri	‘hang from a tree’
bari	‘garden’	bari	‘carry on back’
kiri	‘poor’	kiri	‘to furnish with heddles’
tʃitɯ̃	‘rope’	tʃitɯ̃	‘old’
nini	‘hold’	nini	‘rescue from water’
li	‘necklace’	li	‘heavy’
tɯ	‘oil’	tɯ	‘deep’
tʃu	‘pig’	tʃu	‘speech’
tʃi	‘blood’	tʃi	‘to make’
kɔ	‘go’	kɔ	‘pluck’

2.2. Recording

Five native speakers (male - sp1, sp2, sp3, sp4 and sp5) participated in the production experiment. The speakers were between the ages 18-30 at the time of data collection. The speakers were from Naam Deori village, Jorhat district of Upper Assam. The speakers were bilingual; apart from Deori they were equally fluent in Assamese, the dominant language of the region. The recordings were done in the field in a quiet setting available to the speaker. The target word bearing the tonal contrast was embedded in a fixed sentence frame “I X said” [ã X nina itʃabɛm] where X is the target word. The use of the frame ensured that intonational interference in the target words was uniform and hence predictable. Tone was not marked in the orthography presented to the participants. A method of using pictures of the target word was integrated into our experimental design in order to have one to one correspondence with the actual meaning in an appropriate context. Prior to the task the procedure was explained to the participants. Each word with the carrier sentence along with the relevant picture was randomized and was presented to the participants on the computer screen and they were instructed to read each word four times in the predetermined sentence frame. However, first three of the four repetitions of each target word were used for analysis; the last repetition was discarded to avoid intonational boundary effect. Recording was done using a Shure SM-10 head-mounted microphone connected to a Tascam DR 100 MK II solid state recorder.

2.3. F0 extraction

After the recording, the sound files were transferred to a personal computer for segmentation and annotation. The segmentation and annotation of the target words were done manually using Praat [9]. Since the TBU is the entire phonological word in Deori [11], the pitch property of each tone is expected to spread across the phonological word. Therefore to analyze the pitch properties of each target word the F0 of syllable 1 (henceforth S1) vowel and F0 of syllable 2 (henceforth S2) vowel were taken into consideration. The pitch of the target words were then calculated at 11 successive points from the onset ('start pitch' 0%) of the syllable to the offset ('end point' 100%) of the syllable using a Praat script. Using the same script, duration and intensity values were also measured. The extracted data was then normalized to avoid between speaker variations. The non-normalized F0 values at 10% interval of time were transformed to normalized F0 values using z score normalization method [8][12]. The percentage wise normalized pitch values were averaged across all the three iterations of each word produced by each speaker separately and was written down on a spreadsheet and plotted for graphical representation to observe the distinct pitch contour. The visual inspection of the pitch contours confirms F0 variations across speakers and across words. Fig 1 shows the pitch contour of monosyllabic words and Fig 2 shows the pitch contour of disyllabic words.

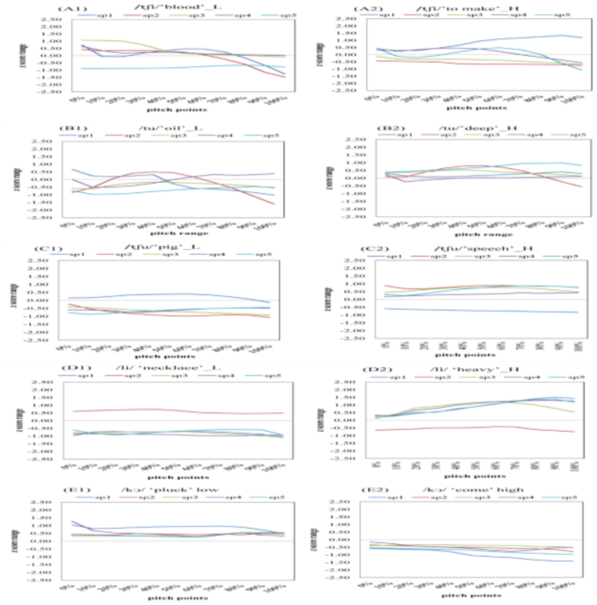

Figure 1: *Average z score normalized pitch contours of each speaker utterance for monosyllabic stems. The left panel shows the low tone words and the right panel shows the high tone words.*

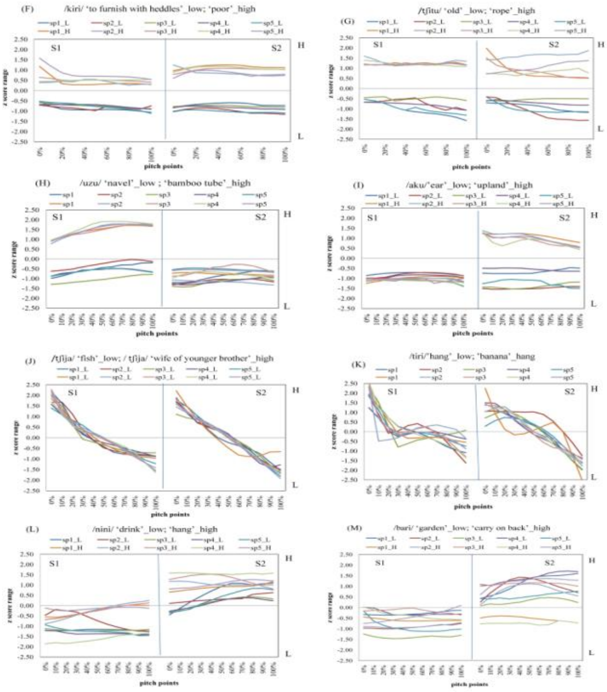

Figure 2: *Average z score normalized pitch contours of each speaker utterance for disyllabic stems words.*

The graphical representation in Fig. 1 and Fig.2 shows that there is considerable F0 variation across words for individual speakers across tones. In order to determine significant effect of tone on the examined words statistical analysis was done the methodology of which is discussed in section 2.4

2.4. Statistical analysis

In order to determine the significant difference between the tonal categories a Univariate Analysis of Variance (ANOVA) was performed. For the statistical analysis average pitch values at 11 successive points from the onset ('start pitch' 0%) of the syllable to the offset ('end point'100%) across the total length of each TBU across speaker was measured. Apart from the F0 values, the duration and the intensity values were also considered to establish the presence of tonal contrast among the speakers. While F0, duration and intensity values of the pitch contour were the dependent variable, tonal categories were the fixed factor. Another ANOVA test was performed to determine the degree of pitch regularity or stability across utterances of particular words. As such a by-word analysis was done to confirm the general observation concerning the variation in pitch patterns across tones for individual speakers. The same method was used for both monosyllabic and disyllabic stems.

3. Results

3.1. Monosyllables

Fig. 1 shows there is a considerable intra-speaker variation in tone realization across individual words. In Fig.1 (A1-A2) shows the pitch contour of *tʃi* where the tonal distinction is maintained by only sp5 (indicated by blue line); sp1 and sp2 completely merge the two tones and sp3 and sp4 reverses the two tonal categories. (B1-B2) shows the pitch contour of *tu* in which the tonal distinction is maintained by sp3 and sp5; sp1, sp2 and sp4 completely merge the two tonal categories (indicated by dark blue, purple and red line respectively). (C1-C2) represents the pitch contour of *tʃu* in which except for sp1 all other speakers' sp2-sp5 maintains tonal distinction; sp1 reverses the two tones. (D1-D2) represents the pitch contour of *li* where all speakers maintain tonal distinction, except for speaker 2 (indicated by red line) who reverses the two tones with a higher (average) F0 value for low tone contour 167.73 Hz and lower (average) F0 value for high tone 129.85 Hz. (E1-E2) represents the pitch contour of *kɔ* in which there is complete tonal reversal with low tone. The underlying low tone *kɔ* is realized as a high tone with higher (average) F0 value 165.21 Hz and the underlying high tone *kɔ* is realized as a low tone with lower (average) F0 value 150.04 Hz. An ANOVA test was conducted to determine whether the monosyllabic words were significantly distinct from each other considering the averaged pitch point values (at 10% interval of time) across all words across speakers. The result shows a significant difference between both the tones ($p<0.05$ at each pitch point) as shown in Table 2.

Table 2: *ANOVA results for effect of tone on F0 for monosyllabic stems (averaged across the pitch point values across each speaker).*

Effect	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Tone	.004	.001	.001	.000	.004	.002	.002	.002	.002	.002	.002

However the overall statistical result across words and across speakers does not justify the general observations concerning the individual speaker variation in pitch patterns. Individual speakers show great variation in pitch realization and statistically all the speakers do not retain tonal distinction in each word. Another ANOVA was performed considering each tonal word for each speaker, with tone as the fixed factor and pitch values (average) at 10% interval of time as the dependant variable. For words like *tʃi* sp1 and sp2 does not have any tonal distinction and it has no significant difference as the *p* value reveals .886 and .712 respectively; sp 3 and sp4 maintains a tonal distinction but they completely reverses the tone categories. It is only sp5 who maintains the tonal distinction ($p<0.05$). For words like *tu* sp1 ($p=.503$), sp2 ($p=.412$) and sp4 ($p=.797$) have no significant tonal distinction. Only sp 3 and sp5 have a tonal distinction with *p* value .002 and .001 respectively. For *tʃu* the statistical result shows a significant distinction of tone for all speakers, but sp1 reverses the two tones. For *li* too there is a significant distinction of tones, but sp 2 reverses the two tones. For *kɔ* all the speakers reverses the two tone categories, but the result shows a significant distinction of tone as there was no tonal overlap. The result further shows that duration and intensity across monosyllabic word types are not statistically significant. An ANOVA test for duration test reveals a *p* value of .382 and for intensity a *p* value of .461. The statistical result thus shows that tone had no significant effect on duration and intensity.

3.2. Disyllables

Fig 2 shows that there is a considerable F0 variation and overlap in disyllabic stems. Tonal distinction in disyllabic word is maintained only in words *kiri* and *tʃitu* across speakers as shown in Fig 2 (F) and (G) respectively. The tonal representation of *uzu* Fig. 2 (H) and aku Fig. 2 (I) is maintained in either of the syllable across speakers. *tʃija* as shown in Fig. 2 (J) and tiri as shown in Fig.2 (K) shows a falling F0 contour without any tonal distinction. *nini* as shown in Fig. 2 (L) has no preserved lexical tonal distinction. *bari* as shown in Fig. 2 (M) shows that tonal distinction is maintained in S2 by only sp1 and sp4 indicated by the orange and the green line respectively. It is also seen that the tonal contours of sp 1and sp 4 merge at final 30% of the pitch measurement point. An ANOVA was done considering the averaged pitch point values (at 10% interval of time) across all words across the speakers. The statistical result (Table 3) shows that the effect of tone is not statistically significant throughout the pitch contour.

Table 3: *ANOVA results for effect of tone on F0 for monosyllabic stems (averaged across the pitch point values across each speaker).*

Tone	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Low*High _S1	.785	.779	.603	.207	.070	.037	.006	.007	.008	.012	.030
Low*High _S2	.121	.031	.027	.116	.254	.009	.010	.003	.005	.003	.003

The difference shows up only in the final four points of S2 out of the eleven points measured. S1 shows no significant difference of tone. Further a by-word analysis was done using ANOVA to determine the effect of tone throughout the pitch contour of each individual word of each individual speaker. The result shows that tone has a significant effect on words *tʃitu* ($p<0.05$) and *kiri* ($p<0.05$) throughout all pitch points in both the syllables across speakers. And for words *uzu* and *aku* tone is significantly different only in S1 and S2 respectively. The result also shows that *bari*, for sp1 and sp4, have a significantly distinct starting average F0 values (Low -165 Hz; High - 192 Hz), but they are more similar at the endpoint measurements, suggesting that they have falling pitch trajectories.

Duration and Intensity values were examined to determine if there is significant effect of duration and intensity. The result showed that duration and intensity had no significant effect on disyllables. An ANOVA test for duration test reveals a *p* value of .075 (S1_low) and .101 (S1_high); .012 (S2_low) and .112 (S2_high) and for intensity a p value of .461 (S1_low) and .021 (S2_high); .091 (S2_low) and .056 (S2_high).

4. Perception Test

After the production experiment a perception test was conducted to reconcile the status of tone among the older and the younger generation speakers. Our hypothesis is that the younger generation speakers tend to misidentify tones compared to the older generation speakers. The purpose of the perception test is to determine whether the listeners would identify the tones as different given the fact that the tonal space between the two tones is smaller. The experiment for perception was designed with natural speech data recorded by an older generation speaker.

4.1. Stimuli

A male native speaker recorded the stimuli for the perception test with CV syllable type as shown in Table 4. The age of the speaker was 52 years at the time of recording. The target word bearing the tonal contrast was embedded in a fixed sentence frame "I X said" [ã X nina itʃabɛm] where X is the target word. The recording was done in an isolated and quiet setting available to the speaker.

Table 4: *The wordlist considered for the perception experiment*

Deori words	Gloss	Deori words	Gloss
[li]	necklace	[li]	heavy
[tʃi]	blood	[tʃi]	to make
[kɔ]	pluck	[kɔ]	come
[tʃu]	pig	[tʃu]	word

4.2. Procedure

An identification task was administered in Praat on a laptop with Sennheiser headphones. All the words were randomized for a total of 200 tokens (8 words*5 repetitions*5 manipulations). The stimuli were played from a computer connected with a headphone. The target stimulus was embedded with three options – the real meaning, the contrastive meaning and 'not sure'. The order of stimulus presentation was randomized. Each of the participants was allowed to listen to one particular sound five times. The speakers listened to a total number of 200 tokens (8 words*5 repetitions*5 manipulations). The speakers were instructed to listen to the stimuli and select any of the three options that they perceive as representing the exact meaning.

4.3. Speakers

Ten native speakers (5 belonging to older and 5 belonging to younger generations) participated in the perception experiment. The participants were all male and they were in the age groups of 18-30 years (younger generation; sp5, sp6, sp7, sp8, sp9) and 45-60 years (older generation; sp10, sp11, sp12, sp13, sp14). The participants were allowed to replay the sound for maximum of five times. The participants had to make choice whether the sound played on the computer were associated with either of the two meanings or 'not sure'.

5. Results

We first discuss the results elicited from older generation speakers. 57.32% of the responses of the older generation speakers could identify the distinct tonal categories. However, sp11 had perceptual difficulty in identifying words like *li* and *tʃi* and opted for the option 'not sure'. A trend of tone reversal was observed in the identification of tone by sp 14 for the word *tʃu*.

For younger generation speakers there was considerable misidentification of tonal categories. The younger generation speakers associated the high tone words with the low tone words and vice versa and failed to associate the exact lexical meaning to the respective tone. For words like *kɔ* and *tʃi* there is a trend of tone reversal and it shows a complete misidentification of tone which conform the production test result. The result shows that apart from the younger generation speakers the older generation speakers also had perceptual difficulty in identifying the tonal categories.

6. Discussion

A look at the statistical analysis done for the targeted words confirm the general observations concerning the great variation in pitch patterns across individual words and also across the speakers overall. The F0 values of the majority of individual words cluster together within each tone, and also there is noticeable variation of F0 within and between (across) tones for individual words. In case of monosyllabic words, the result shows that the speakers still have distinct tone categories but at the same time tone reversal is in process. The two tones are generally separated by only 10 to 20 Hz. Tone is distinct but the tonal space between the two tone categories is very minimal which may indicate that the two tones may completely merge synchronically. For disyllabic words considerable F0 variation was observed across all words and across all speakers. The high degree of individual variation observed in our tone data sheds light on segmental sound changes in progress. The perceptual result too correlates with the idea that perceptual difficulty in identifying the distinct tonal categories can contribute to sound change [7]. The result unveils the initial stages of tonoexodus in progress in Deori. Our data can shed light on how sound change spreads through the lexicon. Our synchronic data of sound change in progress support the idea of lexical diffusion owing to the reason that words across tones were affected differently (Fig. 1 and Fig.2).

The tone system in Deori is the casualty of an egregious contact situation. The overlap and reversal of tone can be attributed to language external factors such as long term bilingual contact with Assamese, a non-tonal language. It has been observed since time immemorial that Tibeto-Burman and Indo-Aryan language have co-existed in the Brahmaputra valley of Assam. For various socio-economic benefits Assamese the dominant language of the region has become the lingua franca for inter and intra group communication for the Deori speakers. In such a process, the use of the native language declines and is restricted to limited domains.

Some of these tonal variations can also be attributed to language-internal factors as well such as the transitional stage from a tone to a stress-accent language. It has been conjectured in a previous study [11] that Deori is in an 'intermediate stage' where tonoexodus is not robust but the language is on its path of emerging as a stress-accent language. Synchronic Deori exhibits an iambic stress pattern [11]. The inherently prominent tonal words may become an anchor for metrical structure facilitating the transition from tone to stress. Consequently, metrical prominence in Deori may have gradually affected the tonal system of the language eventually leading to a completely different stress-accent system. Phonological changes in a language do not affect the whole lexicon at the same time. It is a gradual process and we hypothesize that it is a transitional stage before the whole lexicon undergoes complete transformation.

Certain phonetic variables may also play a role in influencing particular pitch properties of certain words across tones and across speakers. Vowel quality and laryngeal properties may influence slightly different pitch heights within a particular tone. Another possible factor affecting tone merging is word frequency as some words are more prone to sound change than others. This is an area of investigation we intend to explore in future, but as of now we consider the rise of metrical prominence as the more distinguishing factor leading to tonal variations in Deori.

7. References

[1] B.M. Lobanov, "Classification of Russian vowels spoken by different speakers." *The Journal of the Acoustical Society of America*, *49*(2B), 606-608, 1971.

[2] C.Moseley, *Encyclopaedia of the world's endangered languages*. Routledge, 2010.

[3] E. Moulines & J. Laroche, Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech communication*, *16*(2), 175-205, 1995.

[4] F. Jacquesson, *Le deuri: langue tibéto-birmaned'Assam*. vol.88,Peeters Publishers, 2005
[5] J. A. Matisoff, *Handbook of Proto Tibeto-Burman system and philosophy of Sino-Tibetan reconstruction*. 2003.
[6] J.P. Evans, "Is there a Himalayan tone typology." *Senri Ethnological Studies, 75*, 199-221. 2009.
[7] J.J. Ohala, The phonetics of sound change. *Historical linguistics: Problems and perspectives*, 237-278. 1993.
[8] P. Rose, Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech communication*, *6*(4), 343-352, 1987.
[9] P.Boersma, D.Weenink, Praat: doing phonetics by computer [Computer Program]. Version 5.4.08, retrieved 12 August 2016 from http://www.praat.org/.
[10] R.Burling, "Garo," Thurgood Graham and LaPolla, Randy J. (Eds.), *The Sino-Tibetan Languages*, London: Curzon Press, pp.387-400, 2003.
[11] S.Mahanta, I.Dutta, P. Acharyya, "Tonal contrasts in Deori: their existence and word based alignment." *Papers in Historical Phonology*, 2017.
[12] S.F. Disner, Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America*, *67*(1), 253-261, 1980.
[13] U.Goswami, *A Deori Vocabulary and Terminology*. Anundoram Borooah Institue of Language, Art and Culture, 1994.
[14] UNESCO, *Atlas of the World's Languages in Danger* (interactive online edition), 2009.
[15] V. Dybo, S. Nikolayev, and S. Starostin, "A tonological hypothesis on the origin of paradigmatic accent systems." *Estonian papers in phonetics*, *1978*, 16-20, 1978.
[16] W.B. Brown, *An Outline Grammar of the Deori Chutiya Langauge.* Assam secretariat printing Office, 1895.