PHONETIC REALIZATION OF CONTRASTIVELY ASPIRATED AFFRICATES IN NEPALI

G.N. Clements and Rajesh Khatiwada

Laboratoire de Phonétique et Phonologie (UMR 7018, CNRS/Sorbonne-nouvelle, Paris)
clements@idf.ext.jussieu.fr, rajesh.khatiwada@free.fr

ABSTRACT

This paper investigates the phonetic realization of the contrastively aspirated affricates /tsh/ and /dzh/ in Nepali. Static palatography reveals no consistent articulatory difference between these sounds and their nonaspirated counterparts /ts/ and /dz/. However, acoustic examination shows that /tsh/ and /dzh/ differ from /ts/ and /dz/ in several ways depending on the identity of the following vowel (open /a/ vs. palatal /i/). The most prominent and consistent cue to the aspirates, voiceless as well as voiced, is breathy or muffled voice and lowered F0 on the following vowel.

Keywords: Nepali, aspirated affricates, aspiration, breathy voice, [spread glottis].

1. INTRODUCTION

This paper investigates the realization of contrastively aspirated affricates in Nepali, an Indo-Aryan language spoken in Nepal, India and Bhutan [2, 10]. To the best of our knowledge, there are no published phonetic studies of this language. The relevant part of its consonant inventory for our purposes is the following: 1) a voiceless unaspirated stop series /p t ʈ ts k/, 2) a voiced unaspirated stop series /b d ɖ dz g/, 3) a voiceless aspirated stop series /ph th ʈh tsh kh/, and 4) a voiced aspirated ("breathy voiced") stop series /bh dh ɖh dzh gh/. The contrasts of interest for our purposes here are /tsh/ vs. /ts/ and /dzh/ vs. /dz/.

The main goal of this study is to determine whether there is any common acoustic denominator separating the aspirates /tsh/ and /dzh/ from their unaspirated counterparts. Both /tsh/ and /dzh/ are characterized by the feature [+spread glottis] which distinguishes them from nonaspirates [5], and yet in most languages voiceless and voiced aspirates are typically quite different, the first being cued by voiceless aspiration and the second by breathy voice, often accompanied by lowered F0. Affrication introduces a complicating factor, since aspiration can be difficult to distinguish from the frication noise characterizing the affricate release.

Preliminary study of the Nepali affricates /tsh/ and /dzh/ revealed, rather surprisingly, that *both* are distinguished from their nonaspirated counterparts /ts/ and /dz/ by breathy voice and lowered F0 on the following vowel. The purpose of the present study was to confirm this finding by experimental means.

2. METHOD

Here we briefly outline the data and method used in this study.

2.1. Articulatory parameters

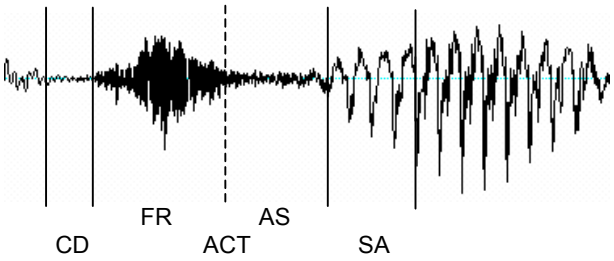
Place of articulation information was determined by classical static palatography [6]. Both palatograms and linguograms were obtained for all test items with the aid of a digital camera. Place of articulation was identified with reference to the anatomical distinctions proposed by Catford [1] and the dentition plan proposed by Firth [3].

2.2. Acoustic and durational parameters

As is well known, the familiar VOT model [7] does not provide for a distinction between plain voiced and voiced aspirated stops such as /d/ vs. /dh/. We have therefore adopted the segmentation model proposed by Mikuteit and Reetz [9] originally designed for the study of the Bengali language. This model recognizes the following phases in stop-vowel sequences, among others: CD (Closure Duration), extending from the onset of the stop closure to the beginning of the release burst, ACT (After Closure Time), extending from the beginning of the release burst to the onset of regular glottal pulsing in the following vowel, and SA (Superimposed Aspiration), the portion of the vowel realized with simultaneous aspiration. SA is characterized by periodic voicing at medium, mainly level amplitude and visible formant structure overlaid by noise excitation.

Figure 1, showing a token of the syllable /tsha/, provides an illustration of these three categories.

Figure 1. Segmented waveform of the Nepali syllable /tsha/, showing CD (Closure Duration), ACT (After Closure Time), and SA (Superimposed Aspiration). The ACT is composed here of successive frication (FR) and aspiration (AS) segments, separated by the dashed line.



In our data, however, there is not usually such a sharp or quantal distinction between the SA segment and the remainder of the vowel, and in our later segmentations we will treat SA as a property of the vowel as a whole.

2.3. Corpus

Two corpora were used. For the articulatory study, we chose four Nepali words containing the phonemes /ts tsh dz dzh/ as follows: *baatsaa* 'commitment', *baatshaa* 'calves', *baadzaa* 'musical instrument', and *baadzhaa* (nonsense word). These words were spoken in the carrier phrase *baabaako X* ('father's X') which contains no coronal or front vowel sounds.

For the acoustic study, we used 8 nonsense words of the form /tsVtsV/, /tshVtshV/, /dzVdzV/, and /dzhVdzhV/ in which V = /a/ or /i/. Each target word X was embedded in the carrier phrase *ma aba* X *bhanchu:* (pause) '*X*' (‘Now I say X: (pause) 'X'), and the full set was recorded 10 times in varying orders by each subject. In these phrases, the first target word occurs in a prefinal context and the second as a citation form.

The role of stress is poorly understood in Nepali. The prefinal target words were produced with approximately equal prominence on both syllables, and only these words are analyzed here.

2.4. Subjects

Four unrelated native speakers of Nepali with no articulatory defects volunteered as subjects for the articulatory experiments. Two were males aged 25 and 57, and two were females aged 39 and 58. All these subjects live in France and use Nepali within the family and frequently in their day-to-day social life.

The acoustic corpus was recorded by two unrelated male native speakers of Nepali aged 36 (M1) and 31 (M2), living in Paris. Neither had any noticeable articulatory idiosyncracies.

2.5. Recording

Recordings were made in a soundproof recording studio at the University of Paris 3 and digitized at a sampling rate of 44100 Hz. Analysis was carried out with Praat (v.4.4.26).

3. RESULTS

3.1. Articulatory results

Table 1 summarizes the articulatory results.

Table 1. Place of articulation of the Nepali affricates /ts tsh dz dzh/ as produced by four speakers. POA= place of articulation; LC = lingual contact; den = dental; alv = alveolar; den-alv = denti-alveolar; post-alv = post-alveolar; ap = apical; lam = laminal.

	ts		tsh		dz		dzh	
	POA	LC	POA	LC	POA	LC	POA	LC
F1	den-alv	lam	den-alv	lam	den-alv	lam	den-alv	lam
F2	alv	lam	alv	lam	alv	lam	alv	lam
M1	post-alv	lam	post-alv	lam	post-alv	lam	post-alv	lam
M2	alv	lam	alv	lam	alv	lam	alv	lam

Affricates were produced as lamino-postalveolar sounds by speaker M1 and as lamino-denti-alveolar or lamino-alveolar sounds by all others. Plain and aspirated sounds showed no place of articulation differences. These results are consistent with the findings of Pokharel [10], who describes the affricates as lamino-alveolar.

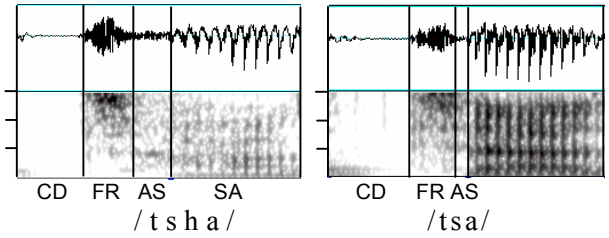
3.2. Acoustic results

In our data, an intervocalic aspirated sound is regularly deaspirated by both speakers; thus, for example, /tshatsha/ is realized [tshatsa]. For this reason, we compare only the first syllables of our test words, where the four-way consonant distinction is maintained.

Representative examples of the first syllables of /tshatsha/ and /tsatsa/ are shown in Fig. 2 (next page). Aspiration is realized in two ways, for both speakers. First, the aspirated consonant /tsh/ is followed by an AS interval averaging 31-32 ms in length, which is typically 3-4 times longer than that of its nonaspirated counterpart /ts/. Second, the vowel following /tsh/ coincides with a SA

segment having the properties described earlier. In many instances, this segment is most prominent at the beginning of the vowel, as in Figs. 1 and 2, but some degree of reduced amplitude and formant attenuation extend over the rest of the vowel.

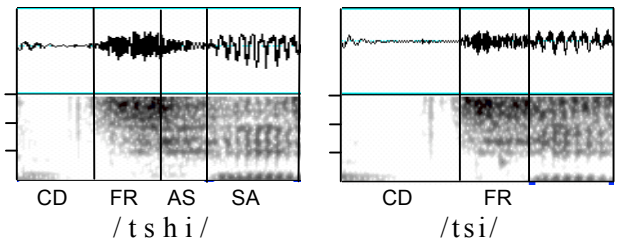
Figure 2. Phonetic realizations of the first syllables of /tshatsha/ and /tsatsa. Waveforms are shown on top and the corresponding spectrograms (range: 0-6 KHz) on the bottom. Speaker: M1.



Auditorily, the vowel of /tsha/ can be described as having a breathy (speaker M1) or muffled (speaker M2) voice quality, contrasting with the ordinary (modal) voicing of /tsa/. A further difference involves F0 values, which were averaged over the first 30 ms of each vowel in the corpus. The F0 of /tsha/ is lower than that of /tsa/ by an average of more than 20 Hz, for both speakers.

Representative examples of the first syllables of /tshitshi/ and /tsitsi/ are shown in Fig. 3.

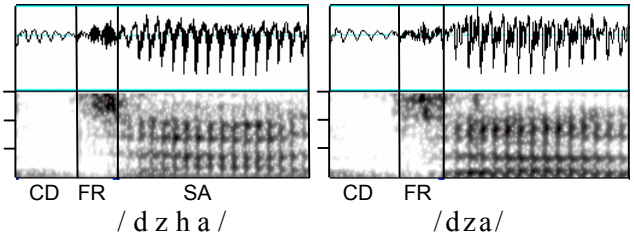
Figure 3. Phonetic realizations of the first syllables of /tshitshi/ and /tsitsi/. Spectrogram range 0-6 KHz. Speaker: M1.



The realization of these affricates is somewhat different from those shown in Fig. 2 due to the influence of the palatal vowel. The most consistent difference between /tshi/ and /tsi/ for both speakers lies in the presence of a robust AS segment in /tshi/, averaging around 30 ms, and its near or total absence in /tsi/. Though this segment is noisy and rather fricative-like, it differs from the FR segment in having major spectral prominences in the 2 - 4.5 KHz region, and it shows a formant pattern similar to that of the following vowel, as is characteristic of aspiration. As for the vowels, both are very short. The vowel of /tshi/, like that of /tsha/, shows reduced amplitude and a relatively indistinct formant pattern. Auditorily, it is dull or muffled in quality, less breathy than the vowel of /tsha/. The vowel of /tsi/, unlike that of /tshi/, shows a strong spectral prominence in the F5 region (4.7-6 KHz), and its waveform shows a noise component. Auditorily, it usually has a fricativized quality (superimposed buzz). The F0 of the initial 30 ms of /tshi/ is marginally lower than that of /tsi/ (average 2 Hz, for both speakers).

Representative examples of the first syllables of /dzhadzha/ and /dzadza/ are shown in Fig. 4.

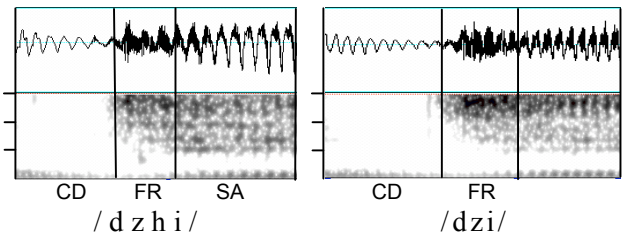
Figure 4. Phonetic realizations of the first syllables of /dzhadzha/ and /dzadza/. Spectrogram range 0-6 KHz. Speaker: M1.



We do not find a voiceless aspiration segment AS in either syllable type. Instead, there is an abrupt transition from FR to the vowel, with an occasional brief amount of vowel lead (in which vowel formants appear under the high-amplitude frication) or vowel lag (in which the frication disappears before clear vowel formants emerge), neither lasting more than two glottal pulses. The main difference between these two syllables is the presence of the SA component in /dzha/, realized auditorily as breathy (M1) or muffled voice (M2) over the full vowel. This difference is accompanied by differences in F0, the vowel of /dzha/ averaging 12 Hz lower for M1 and 22 Hz lower for M2 over its first 30 ms

Representative examples of the first syllables of /dzhidzhi/ and /dzidzi/ are given in Fig. 5.

Figure 5. Phonetic realizations of the first syllables of /dzhidzhi/ and /dzidzi/. Spectrogram range 0-6 KHz. Speaker: M1.



The main difference between these syllables lies in their vowel quality, which is breathy (M1) or muffled (M2) in /dzhi/ but not in /dzi/. The latter instead tends to have the same high-frequency

spectral components and fricativized quality that we found in /tsi/. The two vowels also differ in F0, the initial 30 ms of /dzhi/ being lower than /dzi/ by an average of 13 Hz for M1 and 20 Hz for M2.

4. DISCUSSION

The primary goal of this study was to determine whether there is a common acoustic denominator distinguishing the Nepali aspirated affricates /tsh/ and /dzh/ from their unaspirated counterparts /ts/ and /dz/ in two vowel contexts. We found that the main cue to aspiration, holding of all the contexts examined here, is breathy or muffled voicing extending over the full duration of the following vowel. Several criteria have been suggested for defining breathy voice [4]. One which well separates M1's breathy vowels from his nonbreathy vowels in our data is the greater difference in amplitude between the first and second harmonics in the first few glottal pulses of the vowel. This difference amounts to 9 dB or more for breathy vowels (Table 2):

Table 2. Average differences in decibels between the first and second harmonics (H1-H2) of the vowels in aspirated and unaspirated syllables. Speaker: M1.

aspirated:	H1-H2	unaspirated:	H1-H2
tsha	15.3	tsa	3.5
tshi	9.2	tsi	5.5
dzha	13.8	dza	-0.2
dzhi	9.8	dzi	2.2

However, neither this nor other criteria (H1-A1, H1-A2, H1-A3) clearly separated the vowels of speaker M2. We have used the impressionistic term "muffled" to describe his vowels, which are audibly less breathy than those of speaker M1.

A second cue, also valid for all comparisons, is the lower F0 of the initial 30 msec of vowels following the aspirated consonants. As we have noted, this difference amounts to an average of 12 Hz or more for M1 and of 20 Hz or more for M2 in all comparisons except /tshi/ vs. /tsi/, where it averages only 2 Hz for both speakers.

Other cues to the aspirated vs. unaspirated distinction hold only for some syllables in our data:

1) in /tsha/ and /tshi/, the affrication phase of the affricate is followed by a voiceless aspiration segment AS, often quite long, which contributes to the aspirated auditory quality of the syllable as a whole;
2) in /tsi/ and /dzi/, audible frication noise is superimposed on the vowel; this feature is most likely induced by the presence of the two flanking affricates in the test items, and is not expected to be found in other contexts;
3) in both vowel contexts, the voiced aspirate /dzh/ is about 6 Hz lower in pitch than its unaspirated counterpart /dz/, and tends to be devoiced during the FR segment.

We conclude that breathy or muffled voice (SA), together with F0 lowering, is the most consistent cue to the feature [+spread glottis] in the data studied here. The [-voice] feature of /tsh/ is cued by its voiceless closure duration CD and its voiceless aspiration segment AS.

From a typological point of view, this result is somewhat unusual, as breathy voice and lowered F0 are not commonly observed to be correlates of voiceless aspirates such as /tsh/. However, breathy voice has been found to occur variably on the first few glottal pulses of vowels following voiceless aspirates in English [11, pp. 451-465] and Bengali [9], and both breathy voice and tone "depression" have been reported to occur contrastively on vowels following voiceless aspirated stops and affricates in the Bantu language Ikalanga [8]. Nepali is, then, an unusual but not unique case.

5. REFERENCES

[1] Catford, J.C. 1977. *Fundamental Problems in Phonetics.* Edinburgh: University Press.

[2] Dahal, B.M. 1974. A Description of Nepali Literary and Colloquial. PhD Dissertation, Deccan College, Pune.

[3] Firth, J.R. 1957. Word-palatograms and Articulation. In: *Papers in Linguistics 1934-51.* Oxford, 148-55.

[4] Gordon, M., Ladefoged, P. 2001. Phonation types: a cross-linguistic review. *Journal of Phonetics* 29, 383-406

[5] Halle, M., Stevens, K.N. 1971. A Note on Laryngeal Features. *Quarterly Progress Report* 101, Research Laboratory of Electronics, MIT, Cambridge, Ma., 198-213.

[6] Ladefoged, P. 2003. *Phonetic Data Analysis*, chapter 2. Oxford : Blackwell Publishers.

[7] Lisker, L., Abramson, A.S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384-422.

[8] Mathangwane, J.T. 1999. *Ikalanga Phonetics and Phonology.* Stanford: CSLI.

[9] Mikuteit, S., Reetz H. 2007. Caught in the ACT: The timing of aspiration and voicing in East Bengali. *Language & Speech* 50.2, 249-279.

[10] Pokharel, M.P. 1989. Experimental Analysis of Nepali Sound System. PhD Dissertation, Deccan College, Pune.

[11] Stevens K.N. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.