Analysis of stop consonants in Devanagari alphabet
Kushagra Singh* and Nachiketa Tiwari

*Corresponding author's address: Mechanical Engineering, IIT Kanpur, C-110, Hall 9, Kanpur, 208016, Uttar Pradesh, India, kushagrs@iitk.ac.in

The Devanagari alphabet which is used by several Indian languages including Sanskrit and Hindi has vowels and consonants placed in tabular format which are arranged according to how they originate. A part of this table is a 5 x 5 matrix and comprises of stop consonants, where different rows correspond to velar, palatal, retro-flex, dental and labial stop consonants. In this paper, we have explored patterns that exist between different consonant sounds belonging to different rows and columns of this table. Towards this end, four sound samples from individuals have been recorded, and analysed. Our analysis shows the existence of many interesting relationships which exist between sounds populating different rows and columns of the 5 X 5 matrix. One interesting observation which has been made is that the fundamental differences between 1st and 2nd/3rd or 4th member of each row are essentially the same in all the rows for all the samples, with a few exceptions in the 5th row. Further, the 4th member is found to be the combination of the other two (1-2 and 1-3) variations. In this paper, we present several such interesting relationships. These relationships may be potentially useful in several sound processing algorithms.

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

©2013 Acoustical Society of America [DOI: 10.1121/1.4799571]
Received 22 Jan 2013; published 2 Jun 2013
INTRODUCTION

The Hindi language which is one of the largest spoken languages, uses Devnagari script in which the alphabets are placed in a pattern governed by their phonetics. All consonant sounds are followed by the neutral schwa (ə) by default. Figure 1 shows a part of the Devnagari alphabet chart which consists of stop consonants. Within this 5x5 matrix, the consonants are arranged according to their places of articulation. Here, different rows correspond to velar, palatal, retroflex, dental and bilabial stop consonants respectively. Also, elements belonging to different columns in the same row are placed according to voicing, aspiration and nasality. The unaspirated consonants in this table are also listed in the consonant chart of the IPA.

FIGURE 1. Stop consonants in Hindi language. Different columns correspond to voiceless unaspirated, voiceless aspirated, voiced unaspirated, voiced aspirated and nasal stop consonants for same place of articulation.

Voicing and aspiration, which are the primary factors differentiating consonants in a row of this array have been studied earlier, mainly in terms of VOT (Voice Onset Time) and burst frequency. Lisker and Abramson (1964) have reported VOT values for these consonants (except fifth column and second row) and indicated that these are important distinguishing features in between the consonants in a row. Rami et al. (1999) have done VOT and burst frequency analysis for stops of the first row in this array. Although VOT differences were observed, but burst frequency didn’t fit for a distinguishing feature according to their study. Also, in a spectral pattern study on stop consonant production in children and adults, Nissen and Fox (2009) have used spectral moments to differentiate between /p/, /t/ and /k/. So, frequency domain information can be used to characterize different stops belonging to the same column in this array.

As concluded by Rothenberg (2009) in a study of VOT vs. articulatory modeling of stops, VOT is good as a parameter in speech synthesis and should not be used for analysis of actual speech. In this study, he also pointed out that the voiced aspirated stops in Devnagari, i.e., fourth column stops in the array do not fit into the VOT model. Thus, in this paper VOT has not been focused upon much. Rather, voicing before or during the stop burst has been considered to be different from the voicing which follows the stop burst. Physiologically, the vocal fold vibrations before stop burst and the voicing sound generated after that (which is responsible for the following ‘ə’ sound) have been considered as two different parts in this study, instead of considering both as single ‘voicing’ phenomenon. Also, the spectral properties have been emphasized upon to get a more clear picture of aspiration and voicing. The aim in this study has been to find the fundamental structure of these stop consonants and the relationships between different elements belonging to a single row.

METHOD

Stop consonant voice samples of four persons, from different ages and gender were recorded in a silent environment (Table 1). No past hearing problems were reported by any of them. All the sounds were checked and re-recorded whenever required. Recordings were made using B&K 4192 microphone, sampled at 51.2 kHz and quantized at 24-bit. Each sample was 1 second long and was stored in .wav format (16-bit). To account for the
possible errors emerging from subjectivity involved in drawing conclusions by hearing, all important observations and interpretations were rechecked with other listeners.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>M</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>F</td>
</tr>
</tbody>
</table>

**Procedure: Time Domain Analysis**

Since all stop consonants are followed by the same ‘ə’ vowel sound, the pure consonant parts were separated from the full CV (Consonant Vowel) sound. Different sound segments were played to decide the location of pure consonants. The waveform also played a role in deciding this. This activity was done for each participant and all stop consonants except for the last column which corresponds to nasal stops. As shown in Fig.2, almost all consonants of first four columns can be divided into a pure consonant part, a small transient part and the following vowel (‘ə’) sound.

![Waveform of /kə/ depicting pure consonant sound, transient part and vowel sound](image)

**Procedure: Frequency Domain Analysis**

Due to high sampling rate of 51.2 kHz, it was possible to obtain good resolution in FFT plots. MATLAB® was used to plot signal segments in frequency domain. Logarithmic FFT plots for different segments of the consonant sounds were generated to develop a frequency domain understanding.

**RESULTS**

For all consonants of different participants, the vowel sound part was found to be ‘ə’ indeed. In some cases, it was observed that the transient part was followed by ‘e:jə’ (in first and third elements of the second row) or ‘hə’ (some elements of second and fourth columns). Leaving these exceptions, in other cases interchanging ‘ə’ part in between two stop consonants did not change their sound for same participants. This activity when extended at inter-participant level gave nice results for participants having similar voicing fundamental frequency. So, more emphasis was laid on differences between pure consonants rather than the whole consonant sounds. Also, since the sound samples taken had age and gender based differences, it helped to reach more general conclusions about the relations between consonants in a row/column. It is known that stops arranged in different columns vary in terms of voicing and aspiration. These differences were quantified in time and frequency domains. So, the results are presented in terms of evolution of second, third and fourth column members from the first column. Also, since the VOT values match with those obtained in earlier studies, and as pointed out above, stress has been laid on many other significant characteristics.
**Mahaprana (Second Column)**

The Sanskrit term *mahaprana* is same as aspiration and literally means ‘more air’. The second column stops are called aspirated as the amount of air expelled is more than the first column stops viz. *alpaprana* (meaning less air). It was observed that these aspirated stops have an extra short duration signal located between the pure consonant part and the vowel ‘ə’. This was cross-checked by listening to a waveform without this extra signal, in which case it sounded as the first member. Now onwards, this signal has been called as *mahaprana* in this paper and its structure can be understood from Fig.3. This signal was structurally almost the same for all the aspirated stops of different participants.

![Waveform](image1.png)

**FIGURE 3.** /kʰ/ signal waveform depicting the location of *mahaprana*. If this part is removed and the sound file is played again, it sounds as /k/.

Although VOT is a simple and consistent way of characterizing aspiration, it does not take the shape of *mahaprana* into account. It was tested and confirmed that just by increasing the VOT of unaspirated consonant up to the aspirated level did not help. So, a parameter was needed which could also explain the farness of waveform from zero pressure line which is an important feature of *mahaprana*. The required parameter was chosen as the root square sum of the voltage (which is proportional to pressure) in the *mahaprana* region. Normalization of the unaspirated and aspirated consonants was conducted before calculating this parameter. It takes into account the pressure drops occurring during *mahaprana* which indicate increased air flow. The resulting plot of this parameter for different rows and participants is given below, in Fig.4.

![Plot](image2.png)

**FIGURE 4.** The above plot compares the root square sum voltage values in time domain signal of pure consonant sounds in first and second column of the stop consonant table. It can be noted that the aspirated consonants have much higher value than their unaspirated counterparts in the same row.
Naad (Third Column)

Naad is also a Sanskrit word which would be used in this paper to refer the vocal fold vibration sound before stop burst in voiced consonants. It was observed that what brings a consonant from the first column to the third column is just an addition of a short duration low frequency wave at the starting of the consonant. This is actually the sound of vocal cords vibration before stop burst. The addition was found to occur in two forms—superposition and concatenation. In the former case, the end of naad gets superposed with stop burst sound and in the later case it decays before the stop burst and it is possible to get the first member by removing the naad. These combinations and the structure of naad have been depicted in the figure below, i.e., Fig. 5(a) and 5(b).

FIGURE. 5(a) Waveform of /g/ which is an addition of naad and /k/. End part of the naad is superposed on pure /k/ sound. Sounds as /g/ if started from first vertical line. This is called neer-ksheer7 combination in Sanskrit which means mixture of milk and water, implying superposition of signals.

FIGURE. 5(b) Waveform of /ɖ/ which is an addition of naad and /ʈ/. End part of the naad has decayed before stop burst. Sounds as /ʈ/ if started from first vertical line. The pure /ʈ/ part has low frequency variation also but it is not from the naad. This is called til-tandul7 combination which means mixture of sesame seeds and rice grains, implying decatenation of signals.

In almost all cases, it was possible to transform the first member into third member by transferring the naad and concatenating. Also, in concatenation cases it was possible to extract first member by removing naad from the third member.

Naads of different participants were different in shape but all of them were low frequency waves. However, naads for different voiced consonants for the same speaker were same in shape. In this way, it provides an acoustic print of a person to distinguish him/her from another person. But this doesn’t mean that one person’s naad can not be used to transform another person’s stop. In fact, if the fundamental frequency of two persons’ naad is similar, they can be used interchangeably. For example, /g/ naad of participant 1 was successfully used to convert /k/ of participant 2 into /g/. This was possible because both these had similar voicing frequency. It was observed that this frequency depends on age and gender also, being higher for smaller age (participant 3) and females. The fundamental frequencies of naad part and the following stop burst for different rows have been listed in Table 2. The corresponding column 1 stop burst FFT peak locations have also been provided. It can be observed that in frequency domain as well the frequency corresponding to stop burst is nearly the same for first and third member of a row. Also, the frequency corresponding to naad does not change much as one goes down in a column. Moreover, fundamental frequency of participant 1 and 2 (both 21 years aged males) was found to be almost same. However, their naad waveforms differ significantly in shape. Some sample naad waveforms have been given in Fig.6.
TABLE 2. Fundamental frequencies of naad and stop burst of first and third column stops. These were obtained from FFT magnitude plots of pure consonants. Consonants at odd numbers are from first column and at even numbers are from third column. Frequencies for third column consonants were obtained from pure consonant FFT instead of taking naad and burst FFT separately. Frequencies are in Hz. In the fifth row, i.e., bilabial consonants, the stop burst waveform is not at all periodic. So the burst frequencies don’t match in this case. SB means stop burst.

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Participant 1 naad</th>
<th>Participant 1 SB</th>
<th>Participant 2 naad</th>
<th>Participant 2 SB</th>
<th>Participant 3 naad</th>
<th>Participant 3 SB</th>
<th>Participant 4 naad</th>
<th>Participant 4 SB</th>
</tr>
</thead>
<tbody>
<tr>
<td>/k/</td>
<td>1227</td>
<td></td>
<td>1220</td>
<td></td>
<td>1609</td>
<td></td>
<td>1606</td>
<td></td>
</tr>
<tr>
<td>/g/</td>
<td>125.8</td>
<td>1244</td>
<td>138.5</td>
<td>1247</td>
<td>284</td>
<td>1586</td>
<td>223.5</td>
<td>230.5</td>
</tr>
<tr>
<td>/c/</td>
<td>129</td>
<td>4054</td>
<td>133.4</td>
<td>3511</td>
<td>259</td>
<td>3871</td>
<td>255.7</td>
<td>434.7</td>
</tr>
<tr>
<td>/ɟ/</td>
<td>1645</td>
<td></td>
<td>1456</td>
<td></td>
<td>3476</td>
<td></td>
<td>2120</td>
<td></td>
</tr>
<tr>
<td>/ʈ/</td>
<td>189</td>
<td>1859</td>
<td>133.1</td>
<td>1536</td>
<td>306</td>
<td>3869</td>
<td>242.2</td>
<td>1768</td>
</tr>
<tr>
<td>/ɖ/</td>
<td>1698</td>
<td></td>
<td>1456</td>
<td></td>
<td>2973</td>
<td></td>
<td>1600</td>
<td></td>
</tr>
<tr>
<td>/ṯ/</td>
<td>119</td>
<td>1545</td>
<td>136</td>
<td>2552</td>
<td>297</td>
<td>2035</td>
<td>229.9</td>
<td>1767</td>
</tr>
<tr>
<td>/ḏ/</td>
<td>1080</td>
<td></td>
<td>2351</td>
<td></td>
<td>3755</td>
<td></td>
<td>598</td>
<td></td>
</tr>
<tr>
<td>/p/</td>
<td>118</td>
<td>2357</td>
<td>130</td>
<td>559</td>
<td>308</td>
<td>3774</td>
<td>236.2</td>
<td>1385</td>
</tr>
</tbody>
</table>

FIGURE 6. /ɖ/ naad waveforms of participants 1, 2, 3 and 4 respectively.

Fourth Column

It was observed that the fourth column member can be obtained just by implementing both mahaprana and naad. Thus, the transformations taking first column to second column and first column to third column when combined together, produce the fourth column. The order of individual parts of this voiced aspirated column in time domain is same in all of them and goes like – naad + stop burst + mahaprana + vowel ‘ə’. The naad and mahaprana combinations can be of both superposition and concatenation type. This has been depicted in a sample consonant /ɡ/ in Fig.7. This combination was checked and verified for all the samples.

FIGURE 7. Waveform of /ɡ/ showing naad and mahaprana components.
REFERENCES

1. The International Phonetic Association, The international phonetic alphabet consonant (pulmonic) chart (revised in 2005), http://www.langsci.ucl.ac.uk/ipa/pulmonic.html