Acoustic Study of Hindi Unaspirated Stop Consonants in Consonant-Vowel (CV) Context

R.P.Sharma, I.Khan and O.Farooq

***Abstract*— This paper addresses the acoustic study of the Hindi unaspirated stop consonants in the initial position in a consonant-vowel-consonant (CVC) context with three following vowels / a, i, u/. Eight stop consonant classes of different place of articulations have been taken in initial position of CVC syllables. Acoustic parameters such as voice onset time (VOT), burst duration (BD), burst frequency (BF), formant transition duration (FTD), formant transition frequency (FTF) and formant steady state frequency (SSF) are measured from wave form, frequency spectrum, and spectrogram of CVC syllables. The results show that the VOT duration for all consonant has its lowest value when followed by vowel /a/. BF has its highest value for following vowel /i/ and FTD have its highest and lowest values when followed by vowels /i/ and /u/, respectively, in case of all eight stop consonants. Therefore, the role of following vowel is also important in the acoustic study of Hindi stop consonants.**

Index Terms*— *Acoustic study, Stop consonants

I. Introduction

Acoustic Study of the stop consonants is one of the most challenging tasks in speech recognition due to the dynamic, variable context and speaker-dependent nature of stops. The stop sounds are produced by complex movements in the vocal tract. With the nasal cavity closed, a rapid closure or opening is affected at some points in the oral cavity. Behind the point of closure a pressure is built which is suddenly released with release of closure in vocal tract.

In Hindi, there are 16 stop consonants, while English has only six [1]. The features used for English language may not be useful for Hindi. Thus study of Hindi stop consonants is important in order to understand their time and frequency domain characteristics. This enables us to identify distinguishing features to classify the Hindi stop consonants uniquely. Two parameters required are the voicing during their closure intervals and the place of articulation. The place of articulation classification task is difficult since the acoustic properties of these stop consonants change abruptly during the course of their production. Due to the abrupt nature of stop consonants, traditional statistical methods do not classify them distinctly without the assistance of semantic information. More studies of the acoustic cues for the classification of stop consonants are also needed for the knowledge based approach [1]. The proper selection of cues clearly contributes to the classification performance. Furthermore, the cues should be meaningful in the sense that they should be related to human speech production theory.

R.P.Sharma, Department of Physics, Aligarh Muslim University, Aligarh (INDIA) ,
I. Khan, Department of Physics, Aligarh Muslim University, Aligarh (INDIA)
O. Farooq, Department of Electronics Engineering, Aligarh Muslim University, Aligarh (INDIA)-202002

Several researchers [2-8] have examined the roles played by acoustic cues in the identification of consonants of various categories occupying different positions in a syllable (VC, CV, VCV, CVC, etc.). The stop consonants in initial position of syllables preceding a vowel are cued by various acoustic attributes such as frequency of bursts, onset of the periodic laryngeal vibration or glottal pulsing and the articulatory events associated with the release of the consonant burst and onset frequency of formant transition, etc.

Cooper *et al.*, [2] conducted an experiment to evaluate the role of synthetic burst at specific frequencies placed before synthetic vowels to distinguish among /p, t, k/. Their results shows that the frequency position of burst plus steady-state vowel could serve as a cue, through not necessarily as a completely sufficient one, for the identification of /p, t, k/.

Halle, *et al.*, [3] analyzed the spectral properties of stop bursts containing a number of isolated monosyllabic words. They found that of the three classes of stops associated with different points of articulation, the bilabial stops have a primary concentration of energy in the low frequencies (500-1500 Hz), the postdental stops have either a flat spectrum or one in which the higher frequencies (above 4000 Hz) predominate, and palatal and velar stops have concentration of energy in intermediate frequency regions (500 - 4000 Hz).

Cole and Scott [5] in an experiment with natural CV sounds found that the energy spectrum which accompanies the noise portion burst (release plus aspiration) of a stop consonant in initial position of syllable contains invariant perceptual information. But Dorman *et al.*, [9] found that the burst and transition act in a complementary manner in identifying the initial voiced stops /b, d, g/.

Ohde and Sharf [7] performed experiments with natural stops to evaluate the relative importance of burst and the vowel transition in initial position of CV syllables. They found that burst carries the heaviest load for the identification of unvoiced stops; they also observed that the vowel transition plus steady state vowel is significant to identify unvoiced stops.

In a series of studies Lisker and Abramson [4] have argued that the interval of time measured from the release of an initial stop to the onset of periodicity, denoted as voice onset time (VOT), is the critical acoustic cue for voicing distinctions. In order to do so, the timing of the moment of voice onset has been considered (that is, the timing of the start of vocal cord vibration). They proposed to take the start of the release of the plosive as a reference time. When the value of this reference time is zero, then a moment following the release will have a positive time, and a moment preceding the release will have a negative time. Thus, the VOT is the moment at which the vocal cords start to vibrate, measured in reference to the time of release of the plosive. They also reported that VOT fails to distinguish between voiced unaspirated and aspirated stops.

Winitz *et al.*, [6] found that the duration of VOT was symmetrically altered for English stops and concluded that while aspiration is the primary perceptual cue in the detection of voicing, VOT operates as a relatively unimportant secondary cue. Abramson [10] suggested that VOT is merely one of a large set of interrelated acoustic consequences of variation in the relative timing of glottal and oral gestures. It is often necessary to be able to identify the onset of voicing on the basis of an acoustic analysis alone.

Rami *et al.*, [11] in their study of the VOT and burst frequency of four velar stop consonants in Gujarati found that, voiced stops had significantly higher burst frequencies than unvoiced stops and that there was no significant difference between mean burst frequencies of the aspirated and unaspirated stops. Also the difference in mean VOT as a function of voicing and aspiration were examined. A significant voicing by aspiration effect was found for VOT. The two voiced stops, while not significantly different from each other, had significantly shorter VOTs than unvoiced stops. The aspirated /k^h/ had a significantly longer VOT than the unaspirated /k/.

Banneau *et al.*, [12] reported an experiment on the identification of stops from CVC and CV syllables. The experiment shows that the cues provided by burst onsets under any degree of invariance, are not quite sufficient. First, stop identification can be slightly improved by a foreknowledge of the following vowel. Secondly, the presence of short segment of the following vowel is necessary for perfect stop identification.

Most of these studies are for English and other languages (i.e. two or three category languages). Hindi, an Indo-Aryan language, has four manner categories of stops─voiceless unaspirated, voiced unaspirated, voiceless aspirated and voiced aspirated at four places of articulation─bilabial, dental, post alveolar (retroflex stops), and velar [13]. In Hindi, among the CV syllables that occur in a text about 45% of the syllables belong to the category of stop consonant vowel syllables [14]. Another reason of attention to stops is due to the difficulty in the phoneme classification task [15].

In this paper acoustic study of 8 unaspirated Hindi stop consonants followed by 3 vowel sounds /a, i, u/ is presented. The acoustic study shows that the Hindi stop consonants in initial position of syllables preceding a vowel have various acoustic parameters based on their frequencies and durations. These acoustic parameters are highly affected by the following vowel. Therefore the following vowel may also be plays a very important role in the acoustic study of Hindi stop consonants.

II. MATERIAL

Five speakers, three males and two females, volunteered as speakers for the experiment. The speakers were in age group of 20 to 25 years. None of them had a history of speech, language, or hearing pathology. All speakers had Hindi as their native language and were bilinguals in the sense that they had part of their education through English as their language of instruction.

Eight initial unaspirated consonants, both voiceless and voiced, /p, t, t., k, b, d, d., g/ and 4 final unaspirated voiceless consonants / p, t, t., k / abutted 3 vowel sounds /a, i, u/ to obtain 8 x 3 x 4 = 96 CVC syllables. Some of these syllables were non-sensible. From among these syllables three randomized lists containing 32 words each were prepared to avoid context effects.

Each item was read by the speakers in carrier phrase "/dek^{h}o jʌh CVC hɛ/" in a partially sound treated room and was recorded on a PC with a microphone at a sampling rate of 16 kHz and 16 bits per sample by using “Cool Edit” software. At the time of recording care was taken to keep the distance between microphone and speaker close to 20 cm. Every speaker uttered each list three times. Further, all the CVC syllables were segmented manually from the carrier phrases.

III. PARAMETER MEASUREMENT

To measure the duration and frequency of acoustic features (burst, gap, voice onset time, initial formant transition of vowel, steady state of vowel, final formant transition of vowel) of stop consonants in CVC syllables, waveform and broad-band spectrogram of SFS and Cool Edit software packages were used [8].

A. Voice Onset Time (VOT)

The term Voice Onset Time (VOT) refers to the timing of the beginning of vocal cord vibration in CV sequences relative to the timing of the consonant release as defined earlier. The time difference between release burst of stop consonant and the start of periodic activity (i.e., start of vocal cord vibrations) gives the VOT [4].

B. Burst Frequency and Duration

A speech burst has the form of an impulse and is produced by the release of the closure in the vocal tract. While measuring the duration of the burst, onset of the burst is marked by fixing the points where pattern shows an abrupt change in the overall spectrum after occlusion. The offset of the burst is noted when energy ceases either at a frequency near second formant or higher. In unaspirated stops the offset of the burst is noted as soon as regular glottal pulsing starts. In aspirated stops, the burst from aspirated noise is separated either by the high frequency noise or by a brief period of silence before the onset of aspiration noise. The offset of the burst in unaspirated stops is found easily by observing the absence of acoustic energy in the spectrogram. Burst frequency was measured from the spectra of each consonant. Spectra were obtained, taking the Fast Fourier Transform of the signal to determine the frequencies present. The burst frequency was chosen as the frequency corresponding to the highest amplitude present in the signal spectrum [16].

Duration and formant frequencies of formant transitions (F2 and F3) were measured from the broadband spectrogram. Duration measurements for CVC syllables were made for the burst of initial consonant, CV vowel transition, a combined measurement of the vowel nucleus i.e. steady-state of vowel, the final CV transition, the stop gap closure of the final consonant, and burst of final consonant. The duration of formant transition was selected from the onset of the formant to the steady state of vowel formant. The formant frequency measurements for F2 and F3 were made at the starting point of CV formant transition, i.e. initial formant transition (IFT), steady-state vowel midpoint formant frequency, i.e. steady state frequency (SSF), and at the end point of VC vowel transition, i.e. final formant transition (FFT) and frequency of final burst.

IV. Results and Discussion

Measurements of the acoustic parameters for 480 CVC syllables were done manually. In the following description only the acoustic properties of initial stop consonants in CVC syllables are discussed. Important acoustic parameters for CV syllable are duration of initial burst (BD), frequency of initial burst (BF), VOT duration, duration of second formant transition (FTD), frequency of second formant transition (FTF), and frequency of vowel steady state (SSF). The average values of these parameters with their standard deviations (SDs) are shown in Tables 1 and 2 for unvoiced and voiced stops respectively.

Table 1: Average (mean) values with their standard deviations (S.D.) of various acoustic parameters measured for initial unvoiced stop consonant from CVC syllables.

St op	Following Vowel		VOT (ms)	BF (H	BD (ms	FT D	FTF (Hz)	SSF (Hz)
/p/	/a/	Me	9.2	91	5.8	32.	1484	1630
		S.D	2.9	26	2.8	8.9	143	160
	/i/	Me	11.6	21	7.1	23.	2601	2824
		S.D	4	12	1.8	5.4	419	311
	/u/	Me	19.3	15	6.9	20.	1440	1300
		S.D	9.7	11	2.4	10.	453	449
/t/	/a/	Me	8.8	36	8.1	42.	1841	1648
		S.D	1.8	14	2.1	10.	199	153
	/i/	Me	16.3	40	8.6	25.	2627	2857
		S.D	6.1	12	1.9	8.7	193	290
	/u/	Me	14.1	40	10.2	33.	1597	1165
		S.D	4.7	11	1.8	5.6	179	156
/t./	/a/	Me	8.3	31	7.7	40.	2055	1681
		S.D	1.6	16	1.5	6.9	149	141
	/i/	Me	8.1	38	8.5	21.	2756	2876
		S.D	1.6	10	1.3	6.8	223	302
	/u/	Me	9	21	8.1	32.	1759	1205
		S.D	2.8	10	2.4	8.6	311	85
/k/	/a/	Me	23	16	11	44.	1733	1655
		S.D	6.2	76	2.3	7.8	160	172
	/i/	Me	36	39	10.8	19.	2799	2837
		S.D	11.9	81	3.4	9.9	313	309
	/u/	Me	36.3	16	12.3	23.	1294	1291
		S.D	12.2	15	4.5	10.	312	473

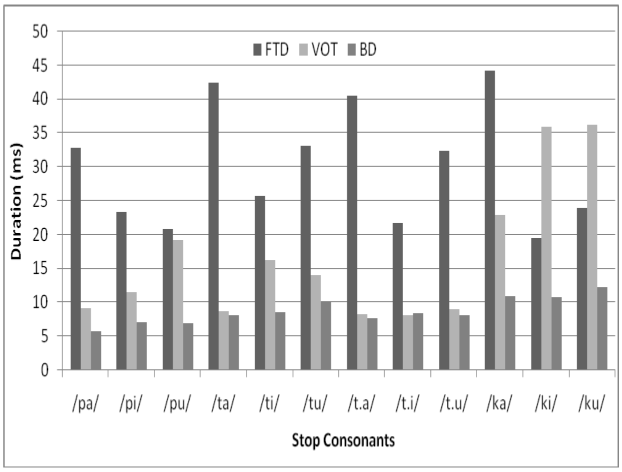

Figure 1: FTD, VOT & BD values of stop consonants /p, t, t., k/ when followed by vowels /a, i, u/.

Table 2: Average (mean) values with their standard deviations (S.D.) of various acoustic parameters measured for initial voiced stop consonant from CVC syllables.

Stop	Following Vowel		VOT (ms)	BF (Hz)	BD (ms)	FTD (ms)	FTF (Hz)	SSF (Hz)
/b/	/a/	Mean	-108.5	100	5.3	27.8	1496	162
		S.D.	16.6	239	2.6	8	149	156
	/i/	Mean	-121.3	215	7	21.9	2570	283
		S.D.	18.7	101	1.2	5.3	371	318
	/u /	Mean	-105.2	117	8.5	21.9	1374	122
		S.D.	25.7	426	3	6.3	435	414
/d/	/a/	Mean	-112.4	415	8.9	45.2	1948	165
		S.D.	23.5	112	2.2	9.9	205	152
	/i/	Mean	-129.9	444	8.3	25.4	2617	282
		S.D.	32.4	832	2.3	7.4	255	293
	/u /	Mean	-121.9	439	8.4	31.8	1728	122
		S.D.	27.9	773	2.1	8.1	182	88
/d./	/a/	Mean	-100.3	332	6.3	41.5	2150	167
		S.D.	12.4	163	3.9	7.8	230	148
	/i/	Mean	-114.2	367	7.1	32.4	2552	257
		S.D.	31.4	106	2.9	16.5	370	569
	/u /	Mean	-115.3	216	8	35.5	1733	124
		S.D.	17.8	901	3.5	10.7	233	85
/g/	/a/	Mean	-91.4	223	8.9	46.2	1820	163
		S.D.	24.7	145	2.6	10.8	148	146
	/i/	Mean	-103.3	413	8.6	27.4	2803	283
		S.D.	23.2	109	2.6	14.8	324	314
	/u /	Mean	-96.5	183	9.1	27.3	1534	147
		S.D.	17	168	2.9	13.3	735	749

The VOT durations for the unvoiced and voiced stop consonants have been grouped as the VOT value for voiced stop consonants is negative and large while for unvoiced stop consonants it is positive and small. For unvoiced stop consonants, the average VOTs for /p, t, t., k/ are 13.4 ms, 13.1 ms, 8.5 ms and 31.8 ms respectively. Thus, the average VOT for different places of articulation is less than 15 ms with the exception of velar /k/ where it is about 30 ms. The VOT is affected by following vowel and is higher for vowel /u/ for all places of articulation. It is lower for all places except for dental for vowel /i/. For vowel /a/ it is distinctly lower for all

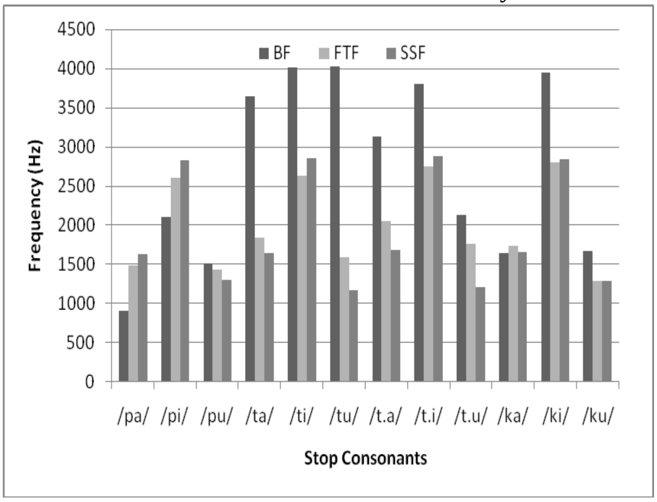

Figure 2: BF, FTF & SSF values of stop consonants /p, t, t., k/ when followed by vowels /a, i, u/.

place of articulations with exception of retroflex. For voiced stop consonants, the average VOTs for /b, d, d. , g/ are -111.7 ms, -121.4 ms, -109.9 ms and -97.1 ms, respectively which shows that VOT is a very important cue for distinction between voiced and unvoiced stop consonants.

Frequencies of second formant transition (FTF) and second formant steady state (SSF) for all stops have maximum values in case of following vowel /i/ and minimum values in case of following vowel /u/. Also BF has highest values for all stop consonants when followed by vowel /i/. Thus FTF, SSF and BF are affected by following vowel for all places of articulations as shown in figures 1-3.

Labial stops (/p/, /b/) have a primary concentration of energy (BF) in the low frequency range (911 – 2153 Hz) with an average of 1477 Hz, whereas average frequency range for dental stops (/t/, /d/) is 3647 to 4448 Hz. For retroflex stops (/t./, /d./) it is found to be from 2126to 3807 Hz, whereas for velar stops (/k/, /g/) frequency range is from 1648 to 4139 Hz. Hence it is concluded that the labial stops have lower burst frequency of about 1500 Hz, and the dental stops have higher burst frequency around 4000 Hz, while the retroflex and velar stops have intermediate ranges of frequency in the nearness of 3000 Hz and 2500 Hz respectively. Also, from the table, it is observed that the burst frequency is affected by the following vowel. It is higher for vowel /i/ for all places of articulation, lower for vowel /a/ in all cases except retroflex stops and also has low values for vowel /u/ in case of dental stop consonants also shown in figures 1&3.

A comparison of the burst frequency with earliest results [3] showed that our values of burst frequency generally fall in the range given by them but for labial stops where they report lower frequency range (500–1500 Hz).In English, VOT for the voiced stops are in general less than 20 ms or even negative, and greater than 20 ms for unvoiced stops [17]. Besides, acoustic study of Hindi retroflex stops is also important.

Khan, *et.al* [18] measured the second formant frequencies of Hindi stop consonants in initial position. They found that average values of second formant frequencies were 1160 Hz, 2500 Hz and 1390 Hz for /pa/, /bi/ and /pu/ respectively. Our values of second formant frequencies also fall in almost similar range as shown in Table 1 and 2.

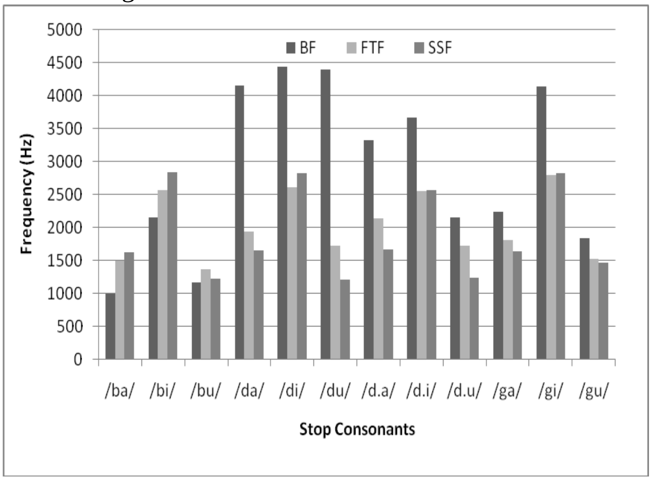

Figure 3: BF, FTF & SSF values of stop consonants /b, d, d., g/ when followed by vowels /a, i, u/.

V. CONCLUSION

Thus the acoustic study shows that the Hindi stop consonants in initial position of syllables preceding a vowel are cued by various acoustic attributes such as frequency of bursts, onset of the periodic laryngeal vibration or glottal pulsing and the articulatory events associated with the release of the consonant burst and onset frequency of formant transition, etc. Therefore, the following vowel plays a very important role in the classification of stop consonants. For Hindi, these cues are different from English and other languages and therefore new feature extraction techniques need to be developed for effective classification of Hindi stop consonants.

VI. ACKNOWLEDGMENT

We are thankful to Mr. S. Hasan Shahid Rizvi for providing valuable help in reshaping this paper.

REFERENCES

[1] A. Suchato, “Classification of stop consonant place of articulation,” Ph.D. dissertation submitted to Massachusetts Institute of Technology, 2004.

[2] S. F. Cooper, P. C. Delattre, and L. J. Gerstman, “Some experiments on the perception of synthetic speech,” J. Acoust. Soc. Am., vol. 24, pp. 597-606, 1952.

[3] M. Halle, G.W. G.Hughes, and J.P.A. Radley, “Acoustic properties of stop consonants,” J. Acoust. Soc. Am., vol. 29, pp. 107-116, 1957.

[4] L. Lisker, and A. Abramson, “A cross study of voicing in initial stops: acoustical measurements,” Word, vol. 20, no. 3, pp. 384, 1964.

[5] R. A. Cole, and B. Scott, “The phantom of the phonemes: Invariant cues for stop consonants,” Perception and Psychophysics, vol. 15, pp. 101-107, 1974.

[6] H. Winitz, C. LaRiviere, and E. Herriman, “Variations in VOT for English initial stops,” J. of Phonetics, vol. 3, pp. 41-52, 1975

[7] R. N. Ohde, and D. J. Sharf, "Order effect of acoustic segments of VC and CV syllables on stop and vowel identification," J. of Speech and Hear. Research, vol. 20, no. 3, pp. 230-237, 1977.
[8] R.P.Sharma, I. Khan, O. Farooq and S.K.Gupta, "Acoustic Study for Recognition of Hindi Stop Consonants in Initial Position of /CVC/ Syllables", J. Acous. Soc. Ind., vol. 37, no. 2, pp. 57-64, 2010.
[9] M. Dorman, M. Studdert-Kennedy, and L. Raphael, "The invariance problems in initial voiced stops: release burst and formant transition as functionally equivalent context dependent cues," Perception and Psychophysics, vol. 22, pp. 109-122, 1977.
[10] A. S. Abramson "Laryngeal timing in consonant distinctions," Phonetica, vol. 34, pp. 295-303, 1977.
[11] Rami, K. Manish, K. Joseph, S. Andrew and P. R. Michael, (1999). "Voice onset time and burst frequencies of four velar stop consonants in Gujarati," J. Acoust. Soc. Am., vol. 106, no. 6, pp. 3736-3738 (L), 1999.
[12] A. Banneau, L. Djezzar, andY. Laprie, "Perception of place of articulation of French stop bursts," J. Acoust. Soc. Am., vol. 100, no. 1, pp. 555-564, 1996.
[13] Ahmed, R. and Agrawal, S.S. (1969). "Significant features in the perception of Hindi consonants," J. Acoust. Soc. Am., 45, 758-763.
[14] C. C. Sekhar, and B. Yegnanarayana, B. "A Constraint satisfaction model for classification of stop consonant–vowel (SCV) utterances," IEEE Transactions on Speech and Audio Processing, vol. 10, no. 7, pp. 472-480, 2002.
[15] O. Farooq, and S. Datta, S., "Mel filter-like admissible wavelet packet structure for speech classification," IEEE Signal Process. Letter, vol. 8(7), pp. 196-198, 2001.
[16] K. N. Stevens, "Acoustic correlates of some phonetic categories," J. Acoust. Soc. Am., vol. 68, pp. 836 – 842, 1980.
[17] V. W. Zue, "Acoustic characteristics of stop consonants: A controlled study," Ph.D. dissertation, Massachusetts Institute of technology, Cambridge, MA, 1976.
[18] I. Khan, S. K. Gupta, and S.H.S. Rizvi, "Formant frequencies of Hindi vowels in /hVd/ and C1VC2 contexts," J. Acoust. Soc. Am., vol. 96, no. 4, pp. 2580-2582, 1994.