

## COMPARATIVE ANALYSIS OF HINDI RETROFLEX AND DENTAL CV SYLLABLES AND THEIR SYNTHESIS

*Rajesh Verma and Puneet Chawla*

Speech Technology Group  
Central Electronics Engineering Research Institute Centre  
CSIR Complex, NPL Campus, Hill Side Road, New Delhi 110 012, India

### Abstract

This paper describes in detail the analysis results of the Hindi Retroflex consonants /t./, /t.<sup>h</sup>/, /d./ & /d.<sup>h</sup>/ and the Dental consonants /t/, /t<sup>h</sup>/, /d/ and /d<sup>h</sup>/ analyzed by using PC based Sensimetrics Speech Station Software. These sounds were analyzed in five long vowel contexts /a/, /i/, /u/, /e/ and /o/ for a very accurate description of their acoustic characteristics/features and the differences between the corresponding cognate sounds in the two classes. Various parameters like duration of closure/voice bar, duration of burst, voice onset time, duration of aspiration, rate of second formant transition and burst frequencies and amplitudes have been studied in details.

The analyzed data was further used to generate the synthetic CV syllables using a cascade/parallel formant synthesizer simulated on a PC. For the synthesis purpose, the source and vocal tract parameters of the synthesizer configuration were selected very carefully. Special attention was paid to the parameters like formant frequencies and their relative amplitudes, which play an important role in making distinction between cognate sounds like /t/ and /t./. The overall burst amplitude also plays a crucial role to make clear distinction in dental and retroflex cognate sounds.

The parametric doc files were modified iteratively, until a satisfactory quality of synthetic sound was obtained. The quality of synthetic speech was evaluated not only by subjective listening but also by matching the spectra of synthetic speech with original speech.

### 1. INTRODUCTION

Every language has its own set of basic sounds, which constitute the language. Hindi speech contains a set of ten pure vowels and 35 consonants, of which about 29 consonants are of frequent usage. These consonants can be conveniently classified according to the manner and place of articulation [1].

The Hindi consonants possess certain special features, which are not so common to European languages and American English. The major discrepancy results from a difference in the number of places at which the stops can be formed by the tip of the tongue making contact with the front portion of the top of the mouth. In fact in Hindi, there is five way distinction in the place of articulation for the stops sounds, compared to English where there is only four way contrast. In English /t/ and /d/ sounds, the tip of tongue comes in contact with alveolar ridge, while in Hindi there is no alveolar sound, rather it has either dental consonants or retroflex. Therefore English /t/ and /d/ are mid-way between Hindi dental and retroflex. So, there is no separate retroflex category in English but they do exist commonly in Hindi, Malayalam and other Indian Languages [2]. Retroflex sounds are made by curling the tip of the tongue up and back so that the underside touches or approaches the back part of the alveolar ridge.

Development and utilization of technology for practical application pre-supposes a large amount of preparatory language specific research. This relates to the studies of acoustic phonetic properties, compilation and labeling of speech data etc. Such studies provide essential knowledge base required for developing voice

I/O systems, like speech recognition and Text to Speech conversion systems [3]. Therefore, detailed study of these sounds are very important for the understanding the nature of these sounds as well as developing the speech input/output systems for Hindi.

## 2. ANALYSIS PROCEDURE

All the consonants were recorded in the form of CVC syllables (where the final consonant was the same as the initial consonant), spoken by single male native Hindi speaker in all five vowel contexts /a/, /i/, /u/, /e/ and /o/ at sampling rate of 12 kHz. These CVC syllables were then broken into CV and VC syllables for the analysis purposes. These CV syllables were then analyzed using a PC equipped with Sensimetrics Speech Station software.

The spectral analysis of these syllables was carried out using digital spectrograms and other techniques like short time FFT and LPC spectra, pitch, formants, waveforms and envelope displayed together. Using the spectrogram most of the important acoustical events were studied in time as well as frequency domain. Duration

and sequences of events in each sound/syllable was noted down using time scale of the spectrogram. Unlike other sounds, which can be described largely in terms of steady-state spectra, stops are transient phonemes and thus are acoustically complex. Hindi CV syllables of the type stop plus vowel consist of at least four phonetic segments, viz. the closure or voicebar, burst, voice onset time (VOT), aspiration and a voiced interval for vowel.

Burst frequencies depend on the place of articulation of the stop as well as the vowel context. The spectral difference among the stop consonants across the places of articulation are primarily reflected in the spectra of the burst & formant transitions of the target vowel. The burst frequencies of the same consonants do show slight variation in different vowel contexts. To get a very good quality of synthetic sounds, burst frequencies and relative amplitudes alongwith overall burst intensity plays a major role. Burst frequencies and relative amplitudes of retroflex and dental sounds in five vowel contexts are shown in Table I through VII respectively.

Table I Comparison of Burst frequencies of Dental and Retroflex sounds in /a/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	500	650	900	500	400	350	700	450
F2 (Hz)	1700	1850	1900	1450	1400	1650	1750	1750
F3 (Hz)	3050	2550	2800	2000	2600	2650	2800	2150
F4 (Hz)	3700	3750	3800	3500	3750	3550	3850	3550
F5 (Hz)	4850	5200	5000	4150	4550	4500	4550	4100

Table II Comparison of Burst frequencies of Dental and Retroflex sounds in /i/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	400	1075	450	350	300	250	400	300
F2 (Hz)	2000	1700	1850	1900	2100	2100	2100	2000
F3 (Hz)	2800	2750	2800	2850	2800	3000	2700	2750
F4 (Hz)	4000	3500	4000	3850	3500	3700	3800	4000
F5 (Hz)	4600	4700	5100	5150	4850	4650	4900	4800

Table III Comparison of Burst frequencies of Dental and Retroflex sounds in /u/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	500	350	475	450	350	250	300	350
F2 (Hz)	1500	1600	1675	1850	1500	1450	1550	1525
F3 (Hz)	2950	2500	2900	2900	2600	2700	2900	3300
F4 (Hz)	4000	3450	3500	3400	3600	3500	3900	4050
F5 (Hz)	4750	4200	4400	4800	4500	4850	4750	5000

Table IV Comparison of Burst frequencies of Dental and Retroflex sounds in /e/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	400	450	400	300	250	275	275	500
F2 (Hz)	1950	1950	1850	2000	1800	1850	1950	2000
F3 (Hz)	2675	2750	2950	2700	2700	2700	2900	2800
F4 (Hz)	3650	3700	4000	3400	3550	3650	3600	3800
F5 (Hz)	4700	4300	4650	4700	4600	4450	4500	4500

Table V Comparison of Burst frequencies of Dental and Retroflex sounds in /o/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	500	500	500	550	250	400	500	300
F2 (Hz)	1450	1900	1900	1950	1625	1650	1500	1550
F3 (Hz)	2900	2450	3200	2800	2650	2350	2900	2700
F4 (Hz)	3400	3400	4000	3350	3900	3500	3900	3700
F5 (Hz)	4200	4600	4650	4200	4600	4600	4650	4900

Table VI Relative burst amplitudes of original retroflex sounds in five vowel context (with relative level difference)

	/a/	/i/	/u/	/e/	/o/
t.	A2F>A3F>A4F>A5F (5, 5, 3)	A4F>A3F>A2F>A5F (7, 10, 3)	A4F>A2F>A5F>A3F (2, 3, 4)	A3F>A4F>A5F>A2F (2, 6, 2)	A4F>A2F>A3F>A5F (7, 5, 5)
t <sup>h</sup>	A4F>A5F>A3F>A2F (0, 5, 1)	A4F>A2F>A3F>A5F (9, 3, 19)	A2F>A3F>A4F>A5F (2, 3, 7)	A4F>A2F>A3F>A5F (3, 2, 6)	A3F>A4F>A2F>A5F (2, 6, 9)
d.	A4F>A2F>A3F>A5F (4, 5, 1)	A3F>A4F>A2F>A5F (1, 9, 1)	A2F>A3F>A4F>A5F (4, 13, 0)	A4F>A3F>A5F>A2F (1, 3, 4)	A4F>A3F>A2F>A5F (1, 0, 16)
d <sup>h</sup>	A4F>A2F>A3F>A5F (1, 1, 1)	A3F>A4F>A2F>A5F (1, 5, 3)	A2F>A3F>A4F>A5F (12, 4, 9)	A4F>A2F>A3F>A5F (2, 3, 4)	A2F>A3F>A4F>A5F (10, 8, 6)

Table VII Relative burst amplitudes of original dental sounds in five vowel context (with relative level difference)

	/a/	/i/	/u/	/e/	/o/
t	A3F>A4F>A2F>A5F (1, 1, 3)	A5F>A4F>A3F>A2F (0, 8, 1)	A4F>A3F>A5F>A2F (7, 6, 7)	A4F>A2F>A5F>A3F (4, 2, 5)	A5F>A4F>A2F>A3F (9, 9, 4)
t <sup>h</sup>	A2F>A4F>A5F>A3F (5, 1, 10)	A3F>A2F>A5F>A4F (6, 4, 7)	A5F>A4F>A2F>A3F (3, 5, 3)	A5F>A4F>A3F>A2F (5, 5, 7)	A4F>A3F>A5F>A2F (2, 6, 2)
d	A5F>A2F>A4F>A3F (1, 4, 4)	A3F>A4F>A5F>A2F (2, 3, 6)	A2F>A4F>A3F>A5F (2, 5, 2)	A4F>A2F>A5F>A3F (5, 0, 4)	A4F>A5F>A3F>A2F (8, 4, 7)
d <sup>h</sup>	A5F>A4F>A3F>A2F (6, 2, 1)	A5F>A4F>A2F>A3F (5, 4, 1)	A4F>A5F>A3F>A2F (5, 0, 0)	A5F>A4F>A3F>A2F (5, 2, 7)	A4F>A3F>A5F>A2F (1, 2, 5)

### 3. SYNTHESIS PROCEDURE

A PC based cascade/parallel formant synthesizer based on Klatt model [4] was used for the synthesis of retroflex and dental sounds. Based on the analyzed data, synthetic CV tokens were created a number of source and tract parameters were adjusted iteratively in order to achieve a close imitation to natural syllables. These include the source parameters such as amplitude of voicing, frication and aspiration (AV, AF, and AH respectively), Open Quotient (OQ) and Spectral Tilt (TL) and vocal tract parameters

like formant frequencies, amplitudes, and their bandwidths.

#### 3.1 Burst Frequencies:

Same formant frequencies of burst (and transitions) can be used for consonants belonging to same place of articulation. For example same values of first four formants can be used for all four consonants /t., t.<sup>h</sup>, d., d.<sup>h</sup>/ in the retroflex group or /t, t<sup>h</sup>, d, d<sup>h</sup>/ in the dental group for a given vowel context. As can be observed from Table VIII through Table XII.

Table VIII Parameters used for generating Burst of the Dental and Retroflex in /a/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	600	700	1000	500	325	350	700	450
F2 (Hz)	1500	1900	1500	1800	1500	1700	1600	1700
F3 (Hz)	3000	2600	2500	2700	2650	2700	2800	2700
F4 (Hz)	3500	3800	3700	3500	3650	3700	3700	3700
F5 (Hz)	4300	5200	4500	4300	4500	4500	4500	4500

Table IX Parameters used for generating Burst of the Dental and Retroflex in /i/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	400	230	400	250	200	250	200	250
F2 (Hz)	2000	2100	2000	1900	2000	2150	1950	2100
F3 (Hz)	3000	2950	3000	2850	2750	3000	2750	2950
F4 (Hz)	4000	3900	3950	4200	4000	3700	3900	3900
F5 (Hz)	4500	5000	4500	4900	4800	5000	5000	5000

Table X Parameters used for generating Burst of the Dental and Retroflex in /u/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	350	250	450	450	280	250	280	250
F2 (Hz)	1600	1700	1600	1700	1600	1700	1600	1700
F3 (Hz)	3100	2900	2750	2900	3100	2900	3100	2900
F4 (Hz)	3900	4000	4100	3800	3900	4000	4000	3800
F5 (Hz)	4700	5000	5000	4500	4700	5000	5000	4500

Table XI Parameters used for generating Burst of the Dental and Retroflex in /e/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	350	400	200	200	200	280	200	200
F2 (Hz)	1900	2000	1900	2100	1900	2100	2100	2100
F3 (Hz)	2650	2650	2800	2800	2550	2600	2900	2800
F4 (Hz)	3450	3650	4250	3900	4000	3600	3600	3900
F5 (Hz)	4300	4700	4700	5000	4650	4600	4800	5000

Table XII Parameters used for generating Burst of the Dental and Retroflex in /o/ context

	t	t.	t <sup>h</sup>	t. <sup>h</sup>	d	d.	d <sup>h</sup>	d. <sup>h</sup>
F1 (Hz)	350	525	450	500	350	400	250	500
F2 (Hz)	1500	1800	1550	1700	1550	1600	1600	1700
F3 (Hz)	2700	2900	2750	2400	3100	3000	2750	2400
F4 (Hz)	4000	3750	3650	3400	3900	3800	4100	3400
F5 (Hz)	4650	4200	3950	4000	4700	4250	5000	4000

### 3.2 Voice Bar

In case of voiced stops, it has been observed that the center frequency of the voicebar varies between 200-300 Hz, and the amplitude of the first resonance is high while all higher resonance are strongly damped. This type of spectral shape is obtained by using a spectral tilt factor (TL). In case of voiced aspirated stops /t.<sup>h</sup>/ and /d.<sup>h</sup>/, a break in the voicebar prior to the aspiration is observed. Duration of Voice Bar is around 80-85 ms in retroflex and 70 to 75 ms in dental sounds depending on the vowel context.

### 3.3 VOT

In case of retroflex stops /t., t.<sup>h</sup>, d., d.<sup>h</sup>/, there is practically no VOT. The voicing starts immediately after the release of burst. While for the dental unaspirated stops /t, d/ there is VOT

of about 5-10ms, while for aspirated dentals / t<sup>h</sup>, d<sup>h</sup> / there is no VOT.

### 3.4 Aspiration

It has been observed that the bandwidths of the formants are narrower (i.e. the formant peaks are better defined) in the case of voiced aspirated sounds /d, d<sup>h</sup>, d., d.<sup>h</sup>/ as compared to unvoiced aspirated sounds /t, t<sup>h</sup>, t., t.<sup>h</sup>/. The formant transitions from burst frequency to the target vowel frequency are part of the aspiration segment. Aspiration duration is about 90-110 ms in case of retroflex sounds, while the duration is 80-100 ms for dental sounds.

The spectrograms of a sample original and synthetic syllable /t.e/ and / t.<sup>h</sup>u / are shown in Figure 1 and Figure 2 respectively.

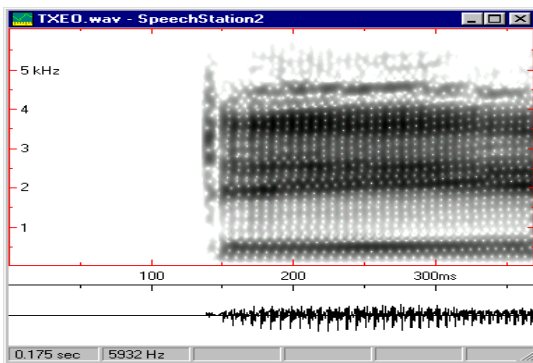


Figure 1.a Spectrogram of original syllable /t.e/

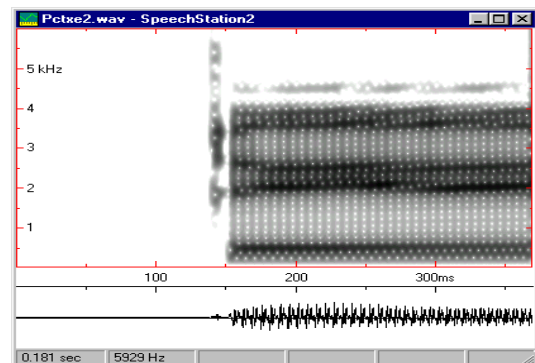


Figure 1.b Spectrogram of synthetic syllable /t.e/

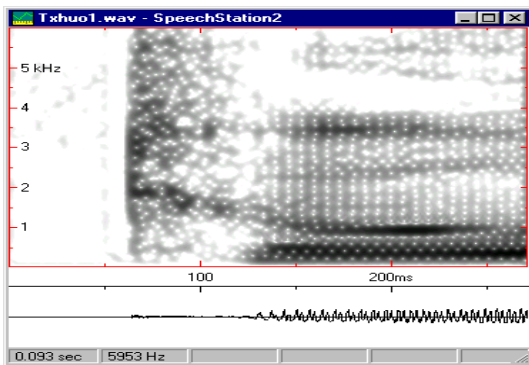


Figure 2.a Spectrogram of original syllable /t.hu/

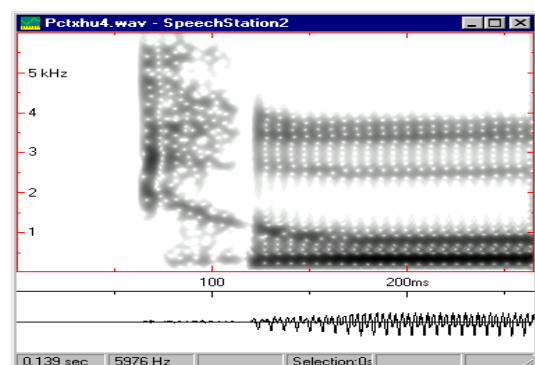


Figure 2.b Spectrogram of synthetic syllable /t.hu/

#### 4. CONCLUSIONS

It has been observed at the burst position that formants three and four contained most of the energy in case of retroflex sounds while formants four and five contained most of the energy in case of dentals. Retroflex sounds have a very strong burst release as compared to the corresponding dental sounds. Also there is general lowering of the third and fourth formants in case of retroflex as compared to dentals. Voice bar frequency is about 10 to 15% lower in case of retroflex as compared to dentals.

The Hindi retroflex and dental sounds have been successfully synthesized using cascade/parallel formant synthesizer. FTP and FTZ, i.e., frequencies of tracheal pole and zero play a very important role in improving the vowel quality. To differentiate between retroflex and dental cognate sounds, the amplitude of burst (AF) plays a crucial role. In general amplitude of burst is higher in case of

retroflex sounds as compared to dental sounds. These synthesis parameters are also used in development of an unlimited vocabulary Text to Speech synthesis system for Hindi.

#### 5. ACKNOWLEDGEMENTS

The authors are grateful to Dr S Ahmad, Director, CEERI, Pilani for his encouragement and useful discussions. They are also thankful to MCIT, Govt. of India, for financial support.

#### 6. REFERENCES

- [1] Agrawal S. S., and Stevens K., "Towards synthesis of Hindi consonants using KLSYN88", Proc. ICSLP-92, 177-180, (1992).
- [2] Peter Ladefoged, "A Course in Phonetics", Harcourt Brace Jovanovich, Inc. , (1975).
- [3] Verma Rajesh et al, "On The Development of Text To Speech System For Hindi", Proc. ICPHS-95, Stockholm, Sweden, vol 2, 1995, p. 354.
- [4] Klatt D. H., " Software for a cascade parallel formant synthesizer", J.Acoust.Soc.Am., 67(3), 971-995, (1980).