

A Proposal for a Phonetic Based Encodng for
Indic scripts
भारतीय लिपियों के लिये स्वनिक कूट का एक
प्रस्ताव

Amitabh Trehan ¹

Supervisors:

Dr. Sanjiva Prasad, Computer Science Department.

Dr. Wagish Shukla, Maths Department

I.I.T. Delhi

¹2000MAC012, M.Tech (C.A.)

Contents

1	Introduction	1
1.1	Outline	2
1.2	The Science of Linguistics and Phonetics	2
1.2.1	Physiology of Speech Production	2
1.2.2	Some definitions	3
2	The cardinal Indian Language phone set	6
3	The Mathematics of Etic Systems	10
4	The encoding scheme	22
4.1	Approaches	22
4.1.1	Feature wise bit encoding (naive)	22
4.1.2	The graph embedding	22
4.1.3	Our approach	23
4.2	The encoding	23
A	Panini's grammar for Sanskrit	32
B	Source Code for helper programs	34
C	The International Phonetic Alphabet	53
D	System and development platform Details	55

List of Tables

2.1	Vowel Chart	7
2.2	Consonant Chart	7
3.1	Definition of Indic Vowels as a collection of features	13
3.2	Definition of Indic Consonants as a collection of features	14
3.2	Definition of Indic Consonants as a collection of features	15
3.3	Etic Distances of Vowels	15
3.4	Etic Distances of Consonants-1	15
3.4	Etic Distances of Consonants-1	16
3.5	Etic Distances of Consonants-2	16
3.5	Etic Distances of Consonants-2	17
3.6	Etic Distances of Consonants-3	18
3.6	Etic Distances of Consonants-3	19
3.7	Etic Distances of Consonants-4	19
3.7	Etic Distances of Consonants-4	20
3.8	Etic Distances of Consonants-5	20
3.8	Etic Distances of Consonants-5	21
4.1	Vowel Encoding	24
4.2	Consonant Encoding	24
4.3	Etic and Hamming Distances of Vowels-I	25
4.4	Etic and Hamming Distances of Vowels-II	25
4.5	Etic and Hamming Distances of Consonants-1	26
4.5	Etic and Hamming Distances of Consonants-1	27
4.6	Etic and Hamming Distances of Consonants-2	27
4.6	Etic and Hamming Distances of Consonants-2	28
4.7	Etic and Hamming Distances of Consonants-3	28
4.7	Etic and Hamming Distances of Consonants-3	29
4.8	Etic and Hamming Distances of Consonants-4	29
4.8	Etic and Hamming Distances of Consonants-4	30
4.9	Etic and Hamming Distances of Consonants-5	30
4.9	Etic and Hamming Distances of Consonants-5	31

List of Figures

1.1	Cross Section of the human vocal tract	3
1.2	The Human Speech Organs	5
2.1	Indian Writing systems charts-I [[9]]	8
2.2	Indian writing Systems-II [[9]]	9
3.1	The Set of Phones	11
3.2	Set of Vowels	12
3.3	Set of Consonants	12
3.4	Partition of Feature Set for Vowel etic system	13
3.5	Partition of Feature Set for Consonant etic system	13
4.1	The Set of Phones	23
C.1	IPA chart courtesy: [[5]].	54

Acknowledgement

I am thankful to my guides, Dr. Sanjiva Prasad and Dr. Wagish Shukla, for suggesting me such an interesting and useful topic to work on, and for guiding and encouraging me throughout. Without their patience with me, and support, I may not have been able to do this modest work.

An area which required particular hard work for this project was the literature and (to some extent) ground survey. Linguistics not being an exact area (My personal feeling), and me being a novice, there were many times I needed help from many sources. I am particularly thankful to Dr. Karel Píška, Institute of Physics, Academy of Sciences, Prague for making to me available photo-copies of [10], during the conference TUG2002. These have proved invaluable. Dr. Rajnish Mishra , Depptt. of Sanskrit, J.N.U. and Dr. Uday Kumar Sharma, Deptt. of Linguistics, Bhagat Singh College sat with me late nights to help compile a chart of phones for Indian languages. For this, I am indebted to them. I also thank Dr. John Plaice, Dr. Anjani Kumar Sinha, and all my friends at the Computer Science department, I.I.T. for their help and suggestions.

There are two organisations I owe a special thanks to. Mahatma Gandhi Antarrashtriya Hindi Vishwavidyalaya, which gave me employment and support to study during my M.Tech here at I.I.T. . I am particularly thankful to the Vice Chancellor, Sh. Ashok Vajpeyi, Dr. Shailendra Singh, Sh. Kaushal Tripathi and all my other colleagues. The TUGIndia (T_EX Users Group, India), which has constantly supported with T_EX related work and has given me the honour of being their board member and by inviting me as a member of the organizing committee for TUG2002. I am particularly thankful to the T_EX guru, Sh. C.V. Radhakrishnan .

Certificate

This is to certify that this theses **Proposal for a phonetics based encoding for Indic scripts** is the bonafide work of **Amitabh Trehan**, *2000MAC012*, M.Tech (Computer Applications) completed under our guidance. This report has not been submitted anywhere else as requirement for completion of any degree.

Sanjiva Prasad
Deptt. of Computer Science

Wagish Shukla
Deptt. of Mathematics

Abstract

Most indic scripts are, what can be termed as phonemic scripts i.e. each character stands for a particular phoneme of the language it is being used for. A *phoneme* is itself, a cover symbol for a collection of sounds (*allophones*). A *phone* is the actual physical sound and *allophones* are the set of *phones* coded by the same *phoneme*. In Indic scripts, the use of allophones is given by simple rules, hence the Indic scripts are rather close to the phonetics (“Study of sound production”) of their sounds. Also, the alphabet (order) of Indian languages has a strong phonetic basis.

We have tried to establish a procedure to exploit this unique mapping, to derive an encoding system based on phonetic properties. The procedure is based on computing a metric called *etic distance* [11] for all the phones and giving them a coding such that the *hamming distance* corresponds closely to the *etic distance* and also tries to respect the alphabetic order i.e. the cultural sensibilities, and, importantly, their phonetic description.

We have collected the standard (or cardinal) phones of Indian scripts (9 scripts used for, at least 13 major Indian languages). These phones have been phonetically classified using the segmental classification method. A mathematical construction called *etic systems* [11] is then used to compute the metric *etic distance*. Then, codes are assigned to each phone to meet the aim stated above. The last step is an ongoing process, which is intuitive in nature and may not be fully realised by the time of writing this report.

Chapter 1

Introduction

Most Indic scripts, *e.g.*, Devanagari for Sanskrit and Hindi, have a phonetic based layout of the alphabet, *i.e.*, the sequence and layout of the standard presentations of the characters have a relation to their sound and place of origin in the human vocal system. There is a unique mapping between characters and sounds. This makes these language alphabets suitable for an encoding mapping characters to their phonetic property. More accurately, these scripts are *phonemic* in nature *i.e.* each character is a phoneme rather than a phone. This simply means that a character may map to a different phone in a different well defined context. These rules have to be encoded for any particular language using a particular script.

Most Indic scripts are “post-syllabic” notations (rather than alphabetic), *i.e.*, the units of notation are renderings of syllables. A reasonable hypothesis is that they are complexes of sounds ending in a vowel. This observation applies across different scripts, and can exploit greater commonality between scripts than can the ISCII or Unicode proposals.

In this project, we have tried to make one such encoding which exploits the closeness of Indic scripts to their phonetics shall be proposed. We have established a procedure which can be used as a basis for this ambitious work, in later projects. The following scripts (and languages) have been covered:

1. Devanagari — Sanskrit, Hindi, Nepali, Marathi, Urdu. [12]
2. Gurmukhi — Punjabi. [14]
3. Gujarati — Gujarati. [18]
4. Bengali — Bengali. [7]
5. Oriya — Oriya. [17]
6. Kannada — Kannada. [13]
7. Telegu — Telegu. [13]
8. Malayalam — Malayalam. [19]
9. Tamil — Tamil. [22]

1.1 Outline

There is a need for a good Hindi (Indian language) encoding at the present moment. Existing schemes are character or rather glyph oriented, and do not adequately address significant cultural issues, particularly transcription. A phonetic based scheme may fill in the blank.

Among other things such an encoding can be used in text-to-speech conversion, speech-to-text conversion and other sound based software and hardware.

The voice interface is particularly important in a context of high illiteracy.

The general methodology used was :

1. Learn the general aspects of phonetics and particular aspects of a subset of Indic languages and their scripts.
2. Collect the total set of Indian phones along with their phonetic description.
3. Find out metrics to determine suitable placements of the character of the alphabet.
4. On the basis of such metrics and phonetic description of the sounds, work out an encoding for the alphabet.

Once such an encoding is established, the following issues could also be investigated (maybe in another project):

- Issues regarding compatibility with and conversions to and from existing encodings (ISCII and Unicode).
- Issues regarding lexicographic orderings and compatibility or otherwise with respect to existing lexicons.
- Issues related to orthography and linear (left-right) and other renderings of the glyphs.

1.2 The Science of Linguistics and Phonetics

1.2.1 Physiology of Speech Production

Many muscles take part in the process of speech production. In brief, however, the sound waves are produced by a complex interaction of

1. An outward flow of air from the lungs,
2. Modifications of the airflow at the larynx (the Adam's apple or "Voice box" in the throat),
3. additional modifications of the airflow by position and movement of the tongue and other anatomical structures of the vocal tract.

A diagram of the human vocal tract is given in Figure 1.1, taken from [3]. The speech organs and the major sounds occurring from them are given in Figure 1.2, taken from [6].

Figure 1.1: Cross Section of the human vocal tract

Figure 1.2: The Human Speech Organs

1.2.2 Some definitions

A useful textbook for beginning study of Linguistics is [1]. The following are some handy definitions taken from this and other sources:

Phonetics Study of how speech sounds are produced (articulated) in the vocal tract (*articulatory phonetics*), and the physical properties of the speech sound waves generated by the vocal tract (*acoustic phonetics*).

Phonology The subfield of linguistics that studies the structure and systematic patterning of sounds in human language. Refers to the abstract rules and principles that govern the distribution of sounds in a language (particular languages or universal properties of natural language sound systems).

The Representation of Speech sounds

Segmental Table A 2-D representation based on the articulation and the organ of production of the sound. This is the traditional method.

Distinctive feature chart A representation based on the internal structure of speech sounds (phones and phonemes) themselves. Popularly proposed by Morris Halle and Noam Chomsky in their 1968 work, *The Sound Pattern Of English (SPE)*, building upon the theory given by Halle and Roman Jakobson.

Consonant A speech sound produced when the speaker either stops or severely constricts the airflow in the vocal tract.

Voicing Vibration in the vocal chord e.g. z-sound in *dogs* versus the s-sound in *cats*.

Stops Sounds produced when the airflow is completely obstructed during speech.

plosive A stop sound in which the release is sudden (explosive), as opposed to an affricate.

Fricatives Sound produced when the airflow is forced through a narrow opening in the vocal tract so that noise produced by friction is created.

Affricates A single but complex sound, beginning as a stop but releasing secondarily into a fricative.

Nasals Voiced oral stops. The airflow and sound energy are channeled into the nasal passages, due to the lowering of the velum.

Liquids a "smooth" sound e.g. /l/ and /r/.

Glides Vowel-like articulations that precede and follow true vowels. also called *semivowels* e.g. /y/ .

Vowels Sounds produced with a relatively open vocal tract which functions as a resonating chamber. produced by shapes of the open, resonating vocal tract and position of several anatomical structures.

- **Muscular Tension** *lax* (or short), *tense* (or long), and *reduced* based on muscular tension and duration. *reduced* vowel are *unstressed* vowels e.g. in English, *schwa*.
- **Shape of Tongue** The shape of the tongue in the mouth is described by the adjectives *low*, *mid*, *high* and *front*, *mid*, *back*.

Diphthong Single vowel sounds beginning in one and ending in another vowel or glide position.

Chapter 2

The cardinal Indian Language phone set

The Indian language phone set was collected through literature research and consultations with experts and native speakers of the language. The phones also had to have proper classification. This involved understanding the science of Phonetics to be able to study the individual classification. we used the segmental classification as it was simpler to handle and data was available for most of the languages in it [10], [16]. There is also scope for investigating how the *Distinctive Feature* classification scheme would hold good here. As yet, *Distinctive Feature* theory is still undergoing refinement, so much clear data is not readily available.

Charts from [9], collated and scanned are given in Figure 2.1 and Figure 2.2. These charts were a good pointer to the classification but were not complete and do not follow the same notations completely. Hence, much work had to be done to complete the classification.

The final classification we made is given in the charts Figure 2.1 and 2.2 .

	Unrounded		Rounded	
	<i>Short</i>	<i>Long</i>	<i>Short</i>	<i>Long</i>
Low Central	अ	आ	ऽ(अ)	आँ
High Front	इ	ई		
High Back			उ	ऊ
<i>Vibrant</i>				
High Front	ऋ	ॠ		
High Back			रु (ऋ)	
<i>Lateral</i>				
High Front	ऌ	ॡ		
SECONDARY VOWELS				
	<i>Short</i>	<i>Long</i>	Diphthong	
UnRounded				
Mid Front	e'	ए		ऐ
Rounded				
Mid Front	ँए			
Mid Back	o'	ओ		औ

Table 2.1: Vowel Chart

	Voiceless		Voiced		Nasals
	<i>Unaspirated</i>	<i>Aspirated</i>	<i>Unaspirated</i>	<i>Aspirated</i>	
Velar	क	ख	ग	घ	ङ
Palatal	च	छ	ज	झ	ञ
Retroflex	ट	ठ	ड	ढ	ण
Dental	त	थ	द	ध	न
Labial	प	फ	ब	भ	म
	<i>Sibilants</i>		<i>Sonorants (voiced)</i>		
			<i>Glide</i>	<i>Liquid</i>	
Palatal	श		य		
Retroflex	ष			र	
Dental	स		ळ	ल	
Labial			व		
Fricatives					
	Voiceless		Voiced		
	<i>Unaspirated</i>	<i>Aspirated</i>	<i>Unaspirated</i>	<i>Aspirated</i>	
Glottal		ह		:	
Velar	क	ख	ग	ङ	
Palatal	tc		ज		
Retroflex			ड	ढ	
Alveolar			र	ण	
Dental			ळ		
Labial		फ			

Table 2.2: Consonant Chart

Figure 2.1: Indian Writing systems charts-I [[9]]

Figure 2.2: Indian writing Systems-II [[9]]

Chapter 3

The Mathematics of Etic Systems

Here we shall follow the description as given in [11].

Linguistically, the phones are described in terms of features. This concept can be explored mathematically. Phonetic systems are sign systems, in the terminology of Charles W. Morris [20]. The unique events in the world are *denotata* of signs (The occurrence of a phone (= a noise)). Any sign system will reduce the variety of experience to a few distinct classes; Each such class of *denotata* to which a sign corresponds is the *designatum* of the sign. In Phonetics, the Designatum is the class of noises classed as equivalent by the linguist in transcription. Finally, the *significatum* of a sign is the set of features which characterize the occurrences of the sign. In Phonetics, this is the Set of features characterizing the class of noises for which the linguist writes the same phone.

Here we shall define an Etic system,.For this, we shall first define a *p.d.f.s.*

Definition 1 Let F be a finite non-empty set $\{a_1, a_2, \dots, a_n\}$, the features, and S another finite non-empty set $\{s_1, s_2, \dots, s_k\}$ the signs or terms, and let there be a function θ with domain S and counterdomain 2^F . The set $\theta(x)$ is the set of features of $x \in S$ or the *significatum* of x . the triple $\langle F, S, \theta \rangle$ composed of F, S , and θ , is called a *preliminary distinctive feature system (p.d.f.s.)*.

When applied to practical data, the features in *p.d.f.s.*'s seem to organise themselves into disjoint subsets of the set of features. The definition of etic systems, this partition is incorporated.

Therefore, an etic system can be defined as:

Definition 2 An etic system is a *p.d.f.s.* $\langle F, S, \theta \rangle$ [Definition 1] whose feature set F is partitioned by a partition P into disjoint subsets $F_1, F_2, \dots, F_n$ such that for each $x \in S$ and each $k = 1, \dots, n$, $\theta(x) \cap F_k$ is either empty or contains exactly one element. This etic system constructed upon $\langle F, S, \theta \rangle$ using partition P is designated $\langle F, S, \theta, P \rangle$. The elements of the partition P can also be called *P-sets*.

In day-to-day application and linguistic study, we often feel some sounds as closer to each other than others. Thus, there are varying degree of similarity between sounds. We need a quantitative measure of this degree of similarity. This is given by the concept of a distance between signs. This is, in certain ways, similar to the distance between two points on a straight line or two points in space. Thus etic distance can be defined as below:

Definition 3 Suppose $\langle F, S, \theta, P \rangle$ is an arbitrary etic system where $P = F_1, F_2, \dots, F_k$. Now consider for each pair $(x, y) \in SXS$ and $i = 1, 2, \dots, k$,

$$D_i(x, y) = \begin{cases} 1 & \text{if } \theta(x) \cap F_i \neq \theta(y) \cap F_i \\ 0 & \text{otherwise} \end{cases}$$

Now let

$$d(x, y) = \sum_{i=1}^k D_i(x, y) \quad (3.1)$$

This function d maps SXS into the set of non-negative integers 0,1,2 ... and is a measure of the feature-distance between the signs x and y . This function $d(x, y)$ can be termed as etic distance.

we also have the concept of a metric etic system which is defined as:

Definition 4 An etic system $\langle F, S, \theta, P \rangle$ in which $d(x, y) = 0$ implies $x = y$ for all x, y in S is a metric etic system.

In phonology, the etic systems are usually metric. The systems we will come across are also metric.

Our set S is given in figure 3.1.

Figure 3.1: The Set of Phones

$$S = \left\{ \begin{array}{l} \text{अ, आ, ॲ, इ, ई, उ, ऊ, ऋ, ॠ, ru} \\ \text{लृ, लॄ, ए, एॄ, ऐ, ओ, ओॄ, औ} \\ \text{ह, ः, क, ख, ग, घ, ङ, च, छ, ज} \\ \text{झ, ञ, ट, ठ, ड, ढ, ण, त, थ} \\ \text{द, ध, न, प, फ, ब, भ, म, य, र} \\ \text{ल,ळ,ळ, व, श, ष, स, क, ख, ग} \\ \text{ज, झ, tc, ड, ण, ढ, फ, } \end{array} \right\}$$

However, due to the different criterions employed to classify consonants and vowels, we take them as two separate etic systems E_v and E_c . Thus, we now have sets S_c (Figure 3.3) and S_v (Figure 3.2).

The feature set F_v is the union of the disjoint partitions $F_{v1} \dots F_{v5}$ and the set of sets $\{F_{v1}, \dots, F_{v5}\}$ constitutes a partition of F . Similarly, for the feature set F_c . These subsets are given in Figure 3.4 and Figure 3.5

$$S_v = \left\{ \begin{array}{l} \text{अ, आ, ॐ, इ, ई, उ, ऊ, ऋ, ॠ, ru} \\ \text{ऌ, ॡ, ए, ए', ऐ, ऐ', ओ, ओ', औ, औ} \end{array} \right\}$$

Figure 3.2: Set of Vowels

$$S_c = \left\{ \begin{array}{l} \text{ह, ः, क, ख, ग, घ, ङ, च, छ, ज} \\ \text{झ, ञ, ट, ठ, ड, ढ, ण, ॠ, त, थ} \\ \text{द, ध, न, प, फ, ब, भ, म, य, र} \\ \text{ल, ळ, ॡ, व, श, ष, स, क, ख, ग} \\ \text{ज, झ, tc, ड, र, ढ, फ} \end{array} \right\}$$

Figure 3.3: Set of Consonants

We then define the phones according to their features. It can be easily verified that both E_v and E_c are etic systems according to Definition 2. The definition of Vowels is given in the Table 3.1 and consonants in Table 3.2.

$$\begin{aligned}
F_{v1} &= \{\text{low, mid, high}\} \\
F_{v2} &= \{\text{front, central, back}\} \\
F_{v3} &= \{\text{unrounded, rounded}\} \\
F_{v4} &= \{\text{Short, Long, Diphthong}\} \\
F_{v5} &= \{\text{vibrant, lateral}\}
\end{aligned}$$

Figure 3.4: Partition of Feature Set for Vowel etic system

$$\begin{aligned}
F_{c1} &= \{\text{stop, fricative, affricate, nasal, liquid, glide}\} \\
F_{c2} &= \{\text{glottal, velar, palatal, retroflex, dental, labial}\} \\
F_{c3} &= \{\text{unvoiced, voiced}\} \\
F_{c4} &= \{\text{unaspirated, aspirated}\}
\end{aligned}$$

Figure 3.5: Partition of Feature Set for Consonant etic system

Phone	Description	F_{v1}	F_{v2}	F_{v3}	F_{v4}	F_{v5}
अ	DEVANAGARI LETTER A	low	central	unrounded	short	
आ	DEVANAGARI LETTER AA	low	central	unrounded	long	
ɔ(अ)	Bengali/Oriya LETTER Rounded A	mid	central	rounded	short	
इ	DEVANAGARI LETTER I	high	front	unrounded	short	
ई	DEVANAGARI LETTER II	high	front	unrounded	long	
उ	DEVANAGARI LETTER U	high	back	rounded	short	
ऊ	DEVANAGARI LETTER UU	high	back	rounded	long	
ऋ	DEVANAGARI LETTER VOCALIC R	high	front	unrounded	short	vibrant
ॠ	DEVANAGARI LETTER VOCALIC RR	high	front	unrounded	long	vibrant
ru(ऋ)	Oriya LETTER VOCALIC R	high	back	rounded	short	vibrant
ऌ	DEVANAGARI LETTER VOCALIC L	high	front	unrounded	shortv	lateral
ॡ	DEVANAGARI LETTER VOCALIC LL	high	front	unrounded	long	lateral
ए	DRAVIDIAN LETTER CANDRA E	mid	front	rounded	short	
e'	DRAVIDIAN LETTER SHORT E	mid	front	unrounded	short	
ए	DEVANAGARI LETTER E	mid	front	unrounded	long	
ऐ	DEVANAGARI LETTER AI		front	unrounded	diphthong	
औ	DEVANAGARI LETTER CANDRA O	low	central	rounded	short	
o'	DRAVIDIAN LETTER SHORT O	mid	back	rounded	short	
ओ	DEVANAGARI LETTER O	mid	back	rounded	long	
औ	DEVANAGARI LETTER AU		back	rounded	diphthong	

Table 3.1: Definition of Indic Vowels as a collection of features

Table 3.2: Definition of Indic Consonants as a collection of features

Phone	Description	F_{c1}	F_{c2}	F_{c3}	F_{c4}
ह	DEVANAGARI LETTER HA	fricative	glottal	unvoiced	aspirated
:	DEV LETTER H (Visarga)	fricative	glottal	voiced	aspirated
क	DEVANAGARI LETTER KA	stop	velar		
ख	DEVANAGARI LETTER KHA	stop	velar		aspirated
ग	DEVANAGARI LETTER GA	stop	velar	voiced	
घ	DEVANAGARI LETTER GHA	stop	velar	voiced	aspirated
ङ	DEVANAGARI LETTER NGA	nasal	velar	voiced	
च	DEVANAGARI LETTER CA	affricate	palatal		
छ	DEVANAGARI LETTER CHA	affricate	palatal		aspirated
ज	DEVANAGARI LETTER JA	affricate	palatal	voiced	
झ	DEVANAGARI LETTER JHA	affricate	palatal	voiced	aspirated
ञ	DEVANAGARI LETTER NYA	nasal	palatal	voiced	
ट	DEVANAGARI LETTER TTA	stop	retroflex		
ठ	DEVANAGARI LETTER TTHA	stop	retroflex		aspirated
ड	DEVANAGARI LETTER DDA	stop	retroflex	voiced	
ढ	DEVANAGARI LETTER DDHA	stop	retroflex	voiced	aspirated
ण	DEVANAGARI LETTER NNA	nasal	retroflex	voiced	
ॢण	DEVANAGARI (Tamil) LETTER NNNA	nasal	alveolar	voiced	
त	DEVANAGARI LETTER TA	stop	dental		
थ	DEVANAGARI LETTER THA	stop	dental		aspirated
द	DEVANAGARI LETTER DA	stop	dental	voiced	
ध	DEVANAGARI LETTER DHA	stop	dental	voiced	aspirated
न	DEVANAGARI LETTER NA	nasal	dental	voiced	
प	DEVANAGARI LETTER PA	stop	labial		
फ	DEVANAGARI LETTER PHA	stop	labial		aspirated
ब	DEVANAGARI LETTER BA	stop	labial	voiced	
भ	DEVANAGARI LETTER BHA	stop	labial	voiced	aspirated
म	DEVANAGARI LETTER MA	nasal	labial	voiced	
य	DEVANAGARI LETTER YA	glide	palatal		
र	DEVANAGARI LETTER RA	liquid	retroflex		
ल	DEVANAGARI LETTER LA	liquid	dental		
ळ	DEVANAGARI(Tamil) LLLA	glide	dental		
ॡ	DEVANAGARI(Marathi) LETTER LLA	fricative	dental		
व	DEVANAGARI LETTER VA	glide	labial		
स	DEVANAGARI LETTER SHA	fricative	palatal		
ष	DEVANAGARI LETTER SSA	fricative	retroflex		
स	DEVANAGARI LETTER SA	fricative	dental		
क्व	DEVANAGARI LETTER QA	fricative	velar	unvoiced	unaspirated
ख्	DEVANAGARI LETTER KHHA	fricative	velar	unvoiced	aspirated
ग	DEVANAGARI LETTER GHHA	fricative	velar	voiced	unaspirated
ज	DEVANAGARI LETTER ZA	fricative	palatal	voiced	unaspirated
झ	DRAVIDIAN LETTER JHA	fricative	palatal	voiced	aspirated
tc	DRAVIDIAN(Telegu) LETTER TC	fricative	palatal	unvoiced	unaspirated
ड	DEVANAGARI LETTER DDDHA	fricative	retroflex	voiced	unaspirated
र	DEVANAGARI(Tamil) LETTER RRA	fricative	alveolar	voiced	unaspirated

Table 3.2: Definition of Indic Consonants as a collection of features

Phone	Description	F_{c1}	F_{c2}	F_{c3}	F_{c4}
ढ	DEVANAGARI LETTER RHA	fricative	dental	voiced	aspirated
फ़	DEVANAGARI LETTER FA	fricative	labial	unvoiced	aspirated

Then, the etic distances are computed according to equation 3.1 and the tables are obtained.

Table 3.3: Etic Distances of Vowels

	अ	आ	ऌ	इ	ई	उ	ऊ	ऋ	ॠ	ru	लृ	लृ	ए	e'	ए	ऐ	ऑ	o'	ओ	औ
अ	0	1	2	2	3	3	4	3	4	4	3	4	3	2	3	3	1	3	4	4
आ	1	0	3	3	2	4	3	4	3	5	4	3	4	3	2	3	2	4	3	4
ऌ	2	3	0	3	4	2	3	4	5	3	4	5	1	2	3	4	1	1	2	3
इ	2	3	3	0	1	2	3	1	2	3	1	2	2	1	2	2	3	3	4	4
ई	3	2	4	1	0	3	2	2	1	4	2	1	3	2	1	2	4	4	3	4
उ	3	4	2	2	3	0	1	3	4	1	3	4	2	3	4	4	2	1	2	2
ऊ	4	3	3	3	2	1	0	4	3	2	4	3	3	4	3	4	3	2	1	2
ऋ	3	4	4	1	2	3	4	0	1	2	1	2	3	2	3	3	4	4	5	5
ॠ	4	3	5	2	1	4	3	1	0	3	2	1	4	3	2	3	5	5	4	5
ru	4	5	3	3	4	1	2	2	3	0	3	4	3	4	5	5	3	2	3	3
लृ	3	4	4	1	2	3	4	1	2	3	0	1	3	2	3	3	4	4	5	5
लृ	4	3	5	2	1	4	3	2	1	4	1	0	4	3	2	3	5	5	4	5
ए	3	4	1	2	3	2	3	3	4	3	3	4	0	1	2	3	2	1	2	3
e'	2	3	2	1	2	3	4	2	3	4	2	3	1	0	1	2	3	2	3	4
ए	3	2	3	2	1	4	3	3	2	5	3	2	2	1	0	2	4	3	2	4
ऐ	3	3	4	2	2	4	4	3	3	5	3	3	3	2	2	0	4	4	4	2
ऑ	1	2	1	3	4	2	3	4	5	3	4	5	2	3	4	4	0	2	3	3
o'	3	4	1	3	4	1	2	4	5	2	4	5	1	2	3	4	2	0	1	2
ओ	4	3	2	4	3	2	1	5	4	3	5	4	2	3	2	4	3	1	0	2
औ	4	4	3	4	4	2	2	5	5	3	5	5	3	4	4	2	3	2	2	0

Table 3.4: Etic Distances of Consonants-1

	ह	:	क	ख	ग	घ	ङ	च	छ	ज
ह	0	1	4	3	4	3	4	4	3	4
:	1	0	4	3	3	2	3	4	3	3
क	4	4	0	1	1	2	2	2	3	3
ख	3	3	1	0	2	1	3	3	2	4
ग	4	3	1	2	0	1	1	3	4	2
घ	3	2	2	1	1	0	2	4	3	3
ङ	4	3	2	3	1	2	0	3	4	2
च	4	4	2	3	3	4	3	0	1	1
छ	3	3	3	2	4	3	4	1	0	2
ज	4	3	3	4	2	3	2	1	2	0
झ	3	2	4	3	3	2	3	2	1	1

Table 3.4: Etic Distances of Consonants-1

	ह	:	क	ख	ग	घ	ङ	च	छ	ज
ज	4	3	3	4	2	3	1	2	3	1
ट	4	4	1	2	2	3	3	2	3	3
ठ	3	3	2	1	3	2	4	3	2	4
ड	4	3	2	3	1	2	2	3	4	2
ढ	3	2	3	2	2	1	3	4	3	3
ण	4	3	3	4	2	3	1	3	4	2
ँण	4	3	3	4	2	3	1	3	4	2
त	4	4	1	2	2	3	3	2	3	3
थ	3	3	2	1	3	2	4	3	2	4
द	4	3	2	3	1	2	2	3	4	2
ध	3	2	3	2	2	1	3	4	3	3
न	4	3	3	4	2	3	1	3	4	2
प	4	4	1	2	2	3	3	2	3	3
फ	3	3	2	1	3	2	4	3	2	4
ब	4	3	2	3	1	2	2	3	4	2
भ	3	2	3	2	2	1	3	4	3	3
म	4	3	3	4	2	3	1	3	4	2
य	4	4	2	3	3	4	3	1	2	2
र	4	4	2	3	3	4	3	2	3	3
ल	4	4	2	3	3	4	3	2	3	3
ळ	4	4	2	3	3	4	3	2	3	3
ळ	3	3	2	3	3	4	3	2	3	3
व	4	4	2	3	3	4	3	2	3	3
श	3	4	3	4	3	4	3	2	3	2
ष	3	4	3	4	3	4	3	3	4	3
स	3	4	3	4	3	4	3	3	4	3
क	2	3	3	3	3	3	3	4	4	4
ख	1	2	3	2	3	2	3	4	3	4
ग	3	2	3	3	2	2	2	4	4	3
ज	3	2	4	4	3	3	3	3	3	2
झ	2	1	4	3	3	2	3	3	2	2
tc	2	3	4	4	4	4	4	3	3	3
ड	3	2	4	4	3	3	3	4	4	3
र	3	2	4	4	3	3	3	4	4	3
ढ़	2	1	4	3	3	2	3	4	3	3
फ़	1	2	4	3	4	3	4	4	3	4

Table 3.5: Etic Distances of Consonants-2

	झ	ज	ट	ठ	ड	ढ	ण	ँण	त	थ
ह	3	4	4	3	4	3	4	4	4	3
:	2	3	4	3	3	2	3	3	4	3
क	4	3	1	2	2	3	3	3	1	2
ख	3	4	2	1	3	2	4	4	2	1

Table 3.5: Etic Distances of Consonants-2

	झ	ञ	ट	ठ	ड	ढ	ण	ण̣	त	थ
ग	3	2	2	3	1	2	2	2	2	3
घ	2	3	3	2	2	1	3	3	3	2
ङ	3	1	3	4	2	3	1	1	3	4
च	2	2	2	3	3	4	3	3	2	3
छ	1	3	3	2	4	3	4	4	3	2
ज	1	1	3	4	2	3	2	2	3	4
झ	0	2	4	3	3	2	3	3	4	3
ञ	2	0	3	4	2	3	1	1	3	4
ट	4	3	0	1	1	2	2	3	1	2
ठ	3	4	1	0	2	1	3	4	2	1
ड	3	2	1	2	0	1	1	2	2	3
ढ	2	3	2	1	1	0	2	3	3	2
ण	3	1	2	3	1	2	0	1	3	4
ण̣	3	1	3	4	2	3	1	0	3	4
त	4	3	1	2	2	3	3	3	0	1
थ	3	4	2	1	3	2	4	4	1	0
द	3	2	2	3	1	2	2	2	1	2
ध	2	3	3	2	2	1	3	3	2	1
न	3	1	3	4	2	3	1	1	2	3
प	4	3	1	2	2	3	3	3	1	2
फ	3	4	2	1	3	2	4	4	2	1
ब	3	2	2	3	1	2	2	2	2	3
भ	2	3	3	2	2	1	3	3	3	2
म	3	1	3	4	2	3	1	1	3	4
य	3	2	2	3	3	4	3	3	2	3
र	4	3	1	2	2	3	2	3	2	3
ल	4	3	2	3	3	4	3	3	1	2
ळ	4	3	2	3	3	4	3	3	1	2
ळ̣	4	3	2	3	3	4	3	3	1	2
व	4	3	2	3	3	4	3	3	2	3
श	3	2	3	4	3	4	3	3	3	4
ष	4	3	2	3	2	3	2	3	3	4
स	4	3	3	4	3	4	3	3	2	3
क्र	4	4	4	4	4	4	4	4	4	4
ख	3	4	4	3	4	3	4	4	4	3
ग	3	3	4	4	3	3	3	3	4	4
ज	2	2	4	4	3	3	3	3	4	4
झ	1	2	4	3	3	2	3	3	4	3
tc	3	3	4	4	4	4	4	4	4	4
ड	3	3	3	3	2	2	2	3	4	4
र	3	3	4	4	3	3	3	2	4	4
ढ	2	3	4	3	3	2	3	3	3	2
फ	3	4	4	3	4	3	4	4	4	3

Table 3.6: Etic Distances of Consonants-3

	द	ध	न	प	फ	ब	भ	म	य	र
ह	4	3	4	4	3	4	3	4	4	4
:	3	2	3	4	3	3	2	3	4	4
क	2	3	3	1	2	2	3	3	2	2
ख	3	2	4	2	1	3	2	4	3	3
ग	1	2	2	2	3	1	2	2	3	3
घ	2	1	3	3	2	2	1	3	4	4
ङ	2	3	1	3	4	2	3	1	3	3
च	3	4	3	2	3	3	4	3	1	2
छ	4	3	4	3	2	4	3	4	2	3
ज	2	3	2	3	4	2	3	2	2	3
झ	3	2	3	4	3	3	2	3	3	4
ञ	2	3	1	3	4	2	3	1	2	3
ट	2	3	3	1	2	2	3	3	2	1
ठ	3	2	4	2	1	3	2	4	3	2
ड	1	2	2	2	3	1	2	2	3	2
ढ	2	1	3	3	2	2	1	3	4	3
ण	2	3	1	3	4	2	3	1	3	2
~ण	2	3	1	3	4	2	3	1	3	3
त	1	2	2	1	2	2	3	3	2	2
थ	2	1	3	2	1	3	2	4	3	3
द	0	1	1	2	3	1	2	2	3	3
ध	1	0	2	3	2	2	1	3	4	4
न	1	2	0	3	4	2	3	1	3	3
प	2	3	3	0	1	1	2	2	2	2
फ	3	2	4	1	0	2	1	3	3	3
ब	1	2	2	1	2	0	1	1	3	3
भ	2	1	3	2	1	1	0	2	4	4
म	2	3	1	2	3	1	2	0	3	3
य	3	4	3	2	3	3	4	3	0	1
र	3	4	3	2	3	3	4	3	1	0
ल	2	3	2	2	3	3	4	3	1	1
ळ	2	3	2	2	3	3	4	3	1	1
ळ	2	3	2	2	3	3	4	3	2	2
व	3	4	3	1	2	2	3	2	1	1
श	3	4	3	3	4	3	4	3	2	3
ष	3	4	3	3	4	3	4	3	3	2
स	2	3	2	3	4	3	4	3	3	3
क्र	4	4	4	4	4	4	4	4	4	4
ख	4	3	4	4	3	4	3	4	4	4
ग	3	3	3	4	4	3	3	3	4	4
ज	3	3	3	4	4	3	3	3	3	4
झ	3	2	3	4	3	3	2	3	3	4
tc	4	4	4	4	4	4	4	4	3	4
ड	3	3	3	4	4	3	3	3	4	3
र	3	3	3	4	4	3	3	3	4	4

Table 3.6: Etic Distances of Consonants-3

	द	ध	न	प	फ	ब	भ	म	य	र
द	2	1	2	4	3	3	2	3	4	4
फ	4	3	4	3	2	3	2	3	4	4

Table 3.7: Etic Distances of Consonants-4

	ल	ळ	ळ	व	श	ष	स	क्र	ख	ग
ह	4	4	3	4	3	3	3	2	1	3
:	4	4	3	4	4	4	4	3	2	2
क	2	2	2	2	3	3	3	3	3	3
ख	3	3	3	3	4	4	4	3	2	3
ग	3	3	3	3	3	3	3	3	3	2
घ	4	4	4	4	4	4	4	3	2	2
ङ	3	3	3	3	3	3	3	3	3	2
च	2	2	2	2	2	3	3	4	4	4
छ	3	3	3	3	3	4	4	4	3	4
ज	3	3	3	3	2	3	3	4	4	3
झ	4	4	4	4	3	4	4	4	3	3
ञ	3	3	3	3	2	3	3	4	4	3
ट	2	2	2	2	3	2	3	4	4	4
ठ	3	3	3	3	4	3	4	4	3	4
ड	3	3	3	3	3	2	3	4	4	3
ढ	4	4	4	4	4	3	4	4	3	3
ण	3	3	3	3	3	2	3	4	4	3
ण	3	3	3	3	3	3	3	4	4	3
त	1	1	1	2	3	3	2	4	4	4
थ	2	2	2	3	4	4	3	4	3	4
द	2	2	2	3	3	3	2	4	4	3
ध	3	3	3	4	4	4	3	4	3	3
न	2	2	2	3	3	3	2	4	4	3
प	2	2	2	1	3	3	3	4	4	4
फ	3	3	3	2	4	4	4	4	3	4
ब	3	3	3	2	3	3	3	4	4	3
भ	4	4	4	3	4	4	4	4	3	3
म	3	3	3	2	3	3	3	4	4	3
य	1	1	2	1	2	3	3	4	4	4
र	1	1	2	1	3	2	3	4	4	4
ल	0	0	1	1	3	3	2	4	4	4
ळ	0	0	1	1	3	3	2	4	4	4
ळ	1	1	0	2	3	3	2	3	3	3
व	1	1	2	0	3	3	3	4	4	4
श	3	3	3	3	0	1	1	3	3	4
ष	3	3	3	3	1	0	1	3	3	4
स	2	2	2	3	1	1	0	3	3	4
क्र	4	4	3	4	3	3	3	0	1	1

Table 3.7: Etic Distances of Consonants-4

	ल	ळ	ळ	व	श	ष	स	क	ख	ग
ख	4	4	3	4	3	3	3	1	0	2
ग	4	4	3	4	4	4	4	1	2	0
ज	4	4	3	4	3	4	4	2	3	1
झ	4	4	3	4	3	4	4	3	2	2
tc	4	4	3	4	2	3	3	1	2	2
ड	4	4	3	4	4	3	4	2	3	1
र	4	4	3	4	4	4	4	2	3	1
ढ़	3	3	2	4	4	4	3	3	2	2
फ़	4	4	3	3	3	3	3	2	1	3

Table 3.8: Etic Distances of Consonants-5

	ज	झ	tc	ड	र	ढ़	फ़
ह	3	2	2	3	3	2	1
:	2	1	3	2	2	1	2
क	4	4	4	4	4	4	4
ख	4	3	4	4	4	3	3
ग	3	3	4	3	3	3	4
घ	3	2	4	3	3	2	3
ङ	3	3	4	3	3	3	4
च	3	3	3	4	4	4	4
छ	3	2	3	4	4	3	3
ज	2	2	3	3	3	3	4
झ	2	1	3	3	3	2	3
ञ	2	2	3	3	3	3	4
ट	4	4	4	3	4	4	4
ठ	4	3	4	3	4	3	3
ड	3	3	4	2	3	3	4
ढ	3	2	4	2	3	2	3
ण	3	3	4	2	3	3	4
~ण	3	3	4	3	2	3	4
त	4	4	4	4	4	3	4
थ	4	3	4	4	4	2	3
द	3	3	4	3	3	2	4
ध	3	2	4	3	3	1	3
न	3	3	4	3	3	2	4
प	4	4	4	4	4	4	3
फ	4	3	4	4	4	3	2
ब	3	3	4	3	3	3	3
भ	3	2	4	3	3	2	2
म	3	3	4	3	3	3	3
य	3	3	3	4	4	4	4
र	4	4	4	3	4	4	4
ल	4	4	4	4	4	3	4

Table 3.8: Etic Distances of Consonants-5

	ज	झ	tc	ड़	र	ढ़	फ़
ळ	4	4	4	4	4	3	4
ळ	3	3	3	3	3	2	3
व	4	4	4	4	4	4	3
श	3	3	2	4	4	4	3
ष	4	4	3	3	4	4	3
स	4	4	3	4	4	3	3
क	2	3	1	2	2	3	2
ख	3	2	2	3	3	2	1
ग	1	2	2	1	1	2	3
ज	0	1	1	1	1	2	3
झ	1	0	2	2	2	1	2
tc	1	2	0	2	2	3	2
ड़	1	2	2	0	1	2	3
र	1	2	2	1	0	2	3
ढ़	2	1	3	2	2	0	2
फ़	3	2	2	3	3	2	0

Chapter 4

The encoding scheme

An encoding scheme has been worked out keeping in view the stated aims.

4.1 Approaches

There are actually three approaches in the scope:

4.1.1 Feature wise bit encoding (naive)

A pretty straightforward approach would be to assign bits to a feature set i.e. the sets in Table 3.1 and Table 3.2 e.g. the set F_{c2} , which has 6 members would require 3-bits to encode. Then, all the phones can be encoded on basis of their unique phonetic description. This approach would take at least 9 bits of encoding (or even more according to the exactness of classification, though it is possible to fold many rows and columns. Also, as such, it offers nothing new and ignores our concepts of distances.

4.1.2 The graph embedding

This is the other extreme. The whole problem can be based on matching the etic distances with hammming distances. This corresponds to the Graph theoretical problem of *Embedding Graphs in Hamming space*. This problem can be stated as given:

Statement 1 *A fully connected undirectional graph ($=G1$) with a large number of nodes (> 50), where all the edges are labelled with distances (values) calculated through a certain method, is to be transformed into $G2$ such that each node is now labelled with a unique address (or Name) and the distance metric for the label of each edge is now a function of the constituent nodes such that the labels of the new edges correspond with the older labels as closely as possible.*

This problem is equivalent to the *hard* problem of *Embedding graphs in euclidean space (or hamming space)*. Though theoretically interesting, this approach may

not offer much from the practical linguistic point of view and is also a difficult approach.

4.1.3 Our approach

Our approach is a mix of the two approaches above and tries to work out a balance between the phonetic description, alphabetic sensibilities and the distance metrics. The features are often combined together across feature sets while maintaining a logical structure. To simplify, we have considered and compared the etic and the hamming distances between neighbouring phones in the order of our alphabetic arrangement.

4.2 The encoding

In our coding scheme, consider the consonants as pure consonants, not as consonant-vowel clusters, though in the charts the full character is depicted. i.e. क (ka) is to be understood as क k.

The general scheme is given in Figure 4.1. The coding scheme for vowels is given in Figure 4.1, and for the consonants in 4.2.

The Comparative charts of etic and hamming distances for the same phone are given in Tables 4.3, 4.5, 4.6, 4.7, 4.8 and 4.9.

Figure 4.1: The Set of Phones

Referring to the Vowel chart, the order of arrangement of rows is from Low-Central, to High-front (2 rows, 010 and 101 -complement of 010), High back (2 rows : notice that 100 is complement of 011), to the secondary vowels m-d-frontal and mid-back. The last two bits are standing for Rounding of tongue and length of utterance (short, long).

Similarly, for consonants, the rows try to respect the place of origin of the phones. The last two bits (i.e. in the columns) stand for voicing and aspiration respectively. Row 1(1000) is for glottals. Row2(1001) is for velar stops and fricatives. Row 3(1010) and 4(1011) accommodate Palatals, whose stops are affricates. Row 5(1100) and 6(1101) are for Retroflex phones. Row 7 (1110) for Dentals and row8 (1111) for Labial stops, fricatives, liquids and nasals. The nasals occupy a column (111) all by themselves.

		00	01	10	11
LC	001	अ	आ	ॐ(अ)	ऑ
HF	010	इ	ई		
HB	011			उ	ऊ
011 HB	100			ru (ऋ)	
010 HF	101	ऋ	ॠ	लृ	लृ
MF	110	e'	ए	ँए	ँए
MB	111	o'	ओ		औ

Table 4.1: Vowel Encoding

	000	001	010	011	100	101	110	111
1000						ह		:
1001	क	ख	ग	घ	क	ख	ग	ङ
1010	च	छ	ज	झ		श	य	ञ
1011	tc		ञ	झ			ळ	
1100	ट	ठ	ड	ढ		ष	र	ण
1101			ड	ढ	र		ळ	ँण
1110	त	थ	द	ध		स	ल	न
1111	प	फ	ब	भ		फ़	व	म

Table 4.2: Consonant Encoding

Table 4.3: Etic and Hamming Distances of Vowels-I

	अ		आ		ॐ		इ		ई		उ		ऊ		ऋ		ॠ		ru		लृ	
अ	0	0	1	1	2	1	2	2	3	3	3	2	4	3	3	1	4	2	4	3	3	2
आ	1	1	0	0	3	2	3	3	2	2	4	3	3	2	4	2	3	1	5	4	4	3
ॐ	2	1	3	2	0	0	3	3	4	4	2	1	3	2	4	2	5	3	3	2	4	1
इ	2	2	3	3	3	3	0	0	1	1	2	2	3	3	1	3	2	4	3	3	1	4
ई	3	3	2	2	4	4	1	1	0	0	3	3	2	2	2	4	1	3	4	4	2	5
उ	3	2	4	3	2	1	2	2	3	3	0	0	1	1	3	3	4	4	1	3	3	2
ऊ	4	3	3	2	3	2	3	3	2	2	1	1	0	0	4	4	3	3	2	4	4	3
ऋ	3	1	4	2	4	2	1	3	2	4	3	3	4	4	0	0	1	1	2	2	1	1
ॠ	4	2	3	1	5	3	2	4	1	3	4	4	3	3	1	1	0	0	3	3	2	2
ru	4	3	5	4	3	2	3	3	4	4	1	3	2	4	2	2	3	3	0	0	3	1
लृ	3	2	4	3	4	1	1	4	2	5	3	2	4	3	1	1	2	2	3	1	0	0
लृ	4	3	3	2	5	2	2	5	1	4	4	3	3	2	2	2	1	1	4	2	1	1
ँए	3	4	4	5	1	3	2	2	3	3	2	2	3	3	3	3	4	4	3	1	3	2
e'	2	3	3	4	2	4	1	1	2	2	3	3	4	4	2	2	3	3	4	2	2	3
ए	3	4	2	3	3	5	2	2	1	1	4	4	3	3	3	3	2	2	5	3	3	4
ँए	3	5	3	4	4	4	2	3	2	2	4	3	4	2	3	4	3	3	5	2	3	3
ऑ	1	2	2	1	1	1	3	4	4	3	2	2	3	1	4	3	5	2	3	3	4	2
o'	3	2	4	3	1	3	3	2	4	3	1	2	2	3	4	1	5	2	2	3	4	2
ओ	4	3	3	2	2	4	4	3	3	2	2	3	1	2	5	2	4	1	3	4	5	3
औ	4	4	4	3	3	3	4	4	4	3	2	2	2	1	5	3	5	2	3	3	5	2

Table 4.4: Etic and Hamming Distances of Vowels-II

	लृ	लृ	ँए	e'	ॠ	ँ	ऑ	o'	ओ	औ
अ	3	2	4	3	3	4	3	5	1	2
आ	4	3	3	2	4	5	3	4	2	1
ॐ	4	1	5	2	1	3	2	4	3	5
इ	1	4	2	5	2	2	1	1	2	2
ई	2	5	1	4	3	3	2	2	1	1
उ	3	2	4	3	2	2	3	3	4	4
ऊ	4	3	3	2	3	3	4	4	3	3

Table 4.5: Etic and Hamming Distances of Consonants-1

	ह	:	क	ख	ग	घ	ङ	च	छ	ज
ह	0	0	1	1	4	3	3	2	4	4
:	1	1	0	0	4	4	3	3	3	3
क	4	3	4	4	0	0	1	1	2	2
ख	3	2	3	3	1	1	0	0	2	2
ग	4	4	3	3	1	1	2	2	0	0
घ	3	3	2	2	2	2	1	1	1	1
ङ	4	2	3	1	2	3	3	2	1	2
च	4	3	4	4	2	2	3	3	3	3
छ	3	2	3	3	3	3	2	2	4	4
ज	4	4	3	3	3	3	4	4	3	3
झ	3	3	2	2	4	4	3	3	3	3
ञ	4	2	3	1	3	5	4	4	2	2
ट	4	3	4	4	1	2	2	3	2	3
ठ	3	2	3	3	2	3	1	2	3	4
ड	4	4	3	3	2	3	3	4	1	2
ढ	3	3	2	2	3	4	2	3	2	3
ण	4	2	3	1	3	5	4	4	2	4
ॠ	4	3	3	2	3	4	4	3	2	3
त	4	4	4	5	1	3	2	4	2	4
थ	3	3	3	4	2	4	1	3	3	5
द	4	5	3	4	2	4	3	5	1	3
ध	3	4	2	3	3	5	2	4	2	4
न	4	3	3	2	3	6	4	5	2	5
प	4	5	4	6	1	2	2	3	2	3
फ	3	4	3	5	2	3	1	2	3	4
ब	4	6	3	5	2	3	3	4	1	2
भ	3	5	2	4	3	4	2	3	2	3
म	4	4	3	3	3	5	4	4	2	4
य	4	3	4	2	2	4	3	5	3	4
र	4	3	4	2	2	4	3	5	3	4
ल	4	4	4	3	2	5	3	6	3	4
ळ	4	4	4	3	2	3	3	4	3	2
ॡ	3	4	3	3	2	3	3	4	3	2
व	4	5	4	4	2	4	3	5	3	4
श	3	1	4	2	3	4	4	3	3	5
ष	3	1	4	2	3	4	4	3	3	5
स	3	2	4	3	3	5	4	4	3	4
क	2	2	3	3	3	1	3	2	3	4
ख	1	1	2	2	3	2	2	1	3	5
ग	3	3	2	2	3	2	3	3	2	3
ज	3	5	2	4	4	2	4	3	3	1
झ	2	4	1	3	4	3	2	3	2	2
tc	2	4	3	5	4	1	4	2	4	2
ड	3	5	2	4	4	2	4	3	4	3
र	3	3	2	4	4	2	4	3	4	5

Table 4.5: Etic and Hamming Distances of Consonants-1

	ह		:		क		ख		ग		घ		ङ		च		छ		ज	
ढ	2	4	1	3	4	3	3	2	3	2	2	1	3	2	4	5	3	4	3	4
फ	1	3	2	4	4	4	3	3	4	5	3	4	4	3	4	4	3	3	4	5

Table 4.6: Etic and Hamming Distances of Consonants-2

	झ	ञ	ट	ठ	ड	ढ	ण	ॠ	त	थ
ह	3 ३	4 २	4 ३	3 २	4 ४	3 ३	4 २	4 ३	4 ४	3 ३
:	2 २	3 1	4 ४	3 ३	3 ३	2 २	3 1	3 २	4 5	3 ४
क	4 ४	3 5	1 २	2 ३	2 ३	3 ४	3 5	3 ४	1 ३	2 ४
ख	3 ३	4 ४	2 ३	1 २	3 ४	2 ३	4 ४	4 ३	2 ४	1 ३
ग	3 ३	2 ४	2 ३	3 ४	1 २	2 ३	2 ४	2 ३	2 ४	3 5
घ	2 २	3 ३	3 ४	2 ३	2 ३	1 २	3 ३	3 २	3 5	2 ४
ङ	3 ३	1 २	3 5	4 ४	2 ४	3 ३	1 २	1 1	3 6	4 5
च	2 २	2 ३	2 २	3 ३	3 ३	4 ४	3 5	3 6	2 1	3 २
छ	1 1	3 २	3 ३	2 २	4 ४	3 ३	4 ४	4 5	3 २	2 1
ज	1 1	1 २	3 ३	4 ४	2 २	3 ३	2 ४	2 5	3 २	4 ३
झ	0 0	2 1	4 ४	3 ३	3 ३	2 २	3 ३	3 ४	4 ३	3 २
ञ	2 1	0 0	3 5	4 ४	2 ४	3 ३	1 २	1 ३	3 ४	4 ३
ट	4 ४	3 5	0 0	1 1	1 1	2 २	2 ३	3 ४	1 1	2 २
ठ	3 ३	4 ४	1 1	0 0	2 २	1 1	3 २	4 ३	2 २	1 1
ड	3 ३	2 ४	1 1	2 २	0 0	1 1	1 २	2 ३	2 २	3 ३
ढ	2 २	3 ३	2 २	1 1	1 1	0 0	2 1	3 २	3 ३	2 २
ण	3 ३	1 २	2 ३	3 २	1 २	2 1	0 0	1 1	3 ४	4 ३
ॠ	3 ४	1 ३	3 ४	4 ३	2 ३	3 २	1 1	0 0	3 5	4 ४
त	4 ३	3 ४	1 1	2 २	2 २	3 ३	3 ४	3 5	0 0	1 1
थ	3 २	4 ३	2 २	1 1	3 ३	2 २	4 ३	4 ४	1 1	0 0
द	3 २	2 ३	2 २	3 ३	1 1	2 २	2 ३	2 ४	1 1	2 २
ध	2 1	3 २	3 ३	2 २	2 २	1 1	3 २	3 ३	2 २	1 1
न	3 २	1 1	3 ४	4 ३	2 ३	3 २	1 1	1 २	2 ३	3 २
प	4 ४	3 5	1 २	2 ३	2 ३	3 ४	3 5	3 ४	1 1	2 २
फ	3 ३	4 ४	2 ३	1 २	3 ४	2 ३	4 ४	4 ३	2 २	1 1
ब	3 ३	2 ४	2 ३	3 ४	1 २	2 ३	2 ४	2 ३	2 २	3 ३
भ	2 २	3 ३	3 ४	2 ३	2 ३	1 २	3 ३	3 २	3 ३	2 २
म	3 ३	1 २	3 5	4 ४	2 ४	3 ३	1 २	1 1	3 ४	4 ३
य	3 २	2 1	2 ४	3 5	3 ३	4 ४	3 ३	3 ४	2 ३	3 ४
र	4 ४	3 ३	1 २	2 ३	2 1	3 २	2 1	3 २	2 ३	3 ४
ल	4 ३	3 २	2 ३	3 ४	3 २	4 ३	3 २	3 ३	1 २	2 ३
ळ	4 ३	3 २	2 5	3 6	3 ४	4 5	3 ४	3 ३	1 ४	2 5
ॡ	4 5	3 ४	2 ३	3 ४	3 २	4 ३	3 २	3 1	1 ४	2 5
व	4 ४	3 ३	2 ४	3 5	3 ३	4 ४	3 ३	3 २	2 ३	3 ४
श	3 २	2 1	3 ४	4 ३	3 5	4 ४	3 ३	3 ४	3 ३	4 २
ष	4 ४	3 ३	2 २	3 1	2 ३	3 २	2 1	3 २	3 ३	4 २
स	4 ३	3 २	3 ३	4 २	3 ४	4 ३	3 २	3 ३	2 २	3 1
क़	4 5	4 ४	4 ३	4 ४	4 ४	4 5	4 ४	4 ३	4 ४	4 5

Table 4.6: Etic and Hamming Distances of Consonants-2

	झ		ञ		ट		ठ		ड		ढ		ण		ॠण		त		थ	
ख	3	4	4	3	4	4	3	3	4	5	3	4	4	3	4	2	4	5	3	4
ग	3	4	3	3	4	4	4	5	3	3	3	4	3	3	3	2	4	5	4	6
ज	2	2	2	3	4	4	4	5	3	3	3	4	3	5	3	4	4	3	4	4
झ	1	1	2	2	4	5	3	4	3	4	2	3	3	4	3	3	4	4	3	3
tc	3	3	3	4	4	3	4	4	4	4	4	5	4	6	4	5	4	2	4	3
ङ	3	4	3	5	3	2	3	3	2	1	2	2	2	3	3	2	4	3	4	4
र	3	6	3	5	4	2	4	3	3	3	3	4	3	3	2	2	4	3	4	4
ढ	2	3	3	4	4	3	3	2	3	2	2	1	3	2	3	1	3	4	2	3
फ़	3	4	4	3	4	4	3	3	4	5	3	4	4	3	4	2	4	3	3	2

Table 4.7: Etic and Hamming Distances of Consonants-3

	द	ध	न	प	फ	ब	भ	म	य	र	
ह	4	5	3	4	4	3	4	5	3	4	3
:	3	4	2	3	3	2	4	6	3	5	2
क	2	4	3	5	3	6	1	2	2	3	4
ख	3	5	2	4	4	5	2	3	1	2	3
ग	1	3	2	4	2	5	2	3	3	4	1
घ	2	4	1	3	3	4	3	4	2	3	2
ङ	2	5	3	4	1	3	3	5	4	4	2
च	3	2	4	3	3	4	2	2	3	3	3
छ	4	3	3	2	4	3	3	3	2	2	4
ज	2	1	3	2	2	3	3	3	4	4	3
झ	3	2	2	1	3	2	4	4	3	3	2
ञ	2	3	3	2	1	1	3	5	4	4	2
ट	2	2	3	3	3	4	1	2	2	3	2
ठ	3	3	2	2	4	3	2	3	1	2	3
ड	1	1	2	2	2	3	2	3	3	4	1
ढ	2	2	1	1	3	2	3	4	2	3	2
ण	2	3	3	2	1	1	3	5	4	4	2
ॠण	2	4	3	3	1	2	3	4	4	3	2
त	1	1	2	2	2	3	1	1	2	2	3
थ	2	2	1	1	3	2	2	2	1	3	3
द	0	0	1	1	1	2	2	2	3	3	2
ध	1	1	0	0	2	1	3	3	2	2	3
न	1	2	2	1	0	0	3	4	4	3	2
प	2	2	3	3	3	4	0	0	1	1	4
फ	3	3	2	2	4	3	1	1	0	0	5
ब	1	1	2	2	2	3	1	1	2	2	3
भ	2	2	1	1	3	2	2	2	1	1	4
म	2	3	3	2	1	1	2	3	3	2	3
य	3	2	4	3	3	2	2	4	3	3	0
र	3	2	4	3	3	2	2	4	3	3	0
ल	2	1	3	2	2	1	2	3	3	4	1

Table 4.7: Etic and Hamming Distances of Consonants-3

	द	ध	न	प	फ	ब	भ	म	य	र										
ळ	2	३	3	4	2	३	2	३	3	4	3	2	4	३	3	2	1	1	1	३
ळ	2	३	3	4	2	३	2	३	3	4	3	2	4	३	3	2	2	३	2	1
व	3	2	4	३	3	2	1	2	2	३	2	1	3	2	2	1	1	2	1	2
श	3	4	4	३	3	2	3	4	4	३	3	5	4	4	3	३	2	2	3	4
ष	3	4	4	३	3	2	3	4	4	३	3	5	4	4	3	३	3	4	2	2
स	2	३	3	2	2	1	3	३	4	2	3	4	4	३	3	2	3	३	3	३
क	4	5	4	6	4	5	4	३	4	4	4	4	4	5	4	4	4	३	4	३
ख	4	6	3	5	4	4	4	4	3	३	4	5	3	4	4	३	4	4	4	4
ग	3	4	3	5	3	4	4	4	4	5	3	३	3	4	3	३	4	2	4	2
ज	3	2	3	३	3	4	4	2	4	३	3	1	3	2	3	३	3	2	4	4
झ	3	३	2	2	3	३	4	३	3	2	3	2	2	1	3	2	3	३	4	5
tc	4	३	4	4	4	5	4	1	4	2	4	2	4	३	4	4	3	३	4	5
ड	3	2	3	३	3	4	4	2	4	३	3	1	3	2	3	३	4	4	3	2
र	3	4	3	5	3	4	4	2	4	३	3	३	3	4	3	३	4	4	4	2
ढ	2	३	1	2	2	३	4	३	3	2	3	2	2	1	3	2	4	5	4	३
फ	4	4	3	३	4	2	3	2	2	1	3	३	2	2	3	1	4	4	4	4

Table 4.8: Etic and Hamming Distances of Consonants-4

	ल		ळ		ळ		व		श		ष		स		क		ख		ग	
ह	4	4	4	4	3	4	4	5	3	1	3	1	3	2	2	2	1	1	3	3
:	4	3	4	3	3	3	4	4	4	2	4	2	4	3	3	3	2	2	2	2
क	2	5	2	3	2	3	2	4	3	4	3	4	3	5	3	1	3	2	3	2
ख	3	6	3	4	3	4	3	5	4	3	4	3	4	4	3	2	2	1	3	3
ग	3	4	3	2	3	2	3	3	3	5	3	5	3	6	3	2	3	3	2	1
घ	4	5	4	3	4	3	4	4	4	4	4	4	4	5	3	3	2	2	2	2
ङ	3	4	3	2	3	2	3	3	3	3	3	3	3	4	3	2	3	1	2	1
च	2	3	2	3	2	5	2	4	2	2	3	4	3	3	4	3	4	4	4	4
छ	3	4	3	4	3	6	3	5	3	1	4	3	4	2	4	4	3	3	4	5
ज	3	2	3	2	3	4	3	3	2	3	3	5	3	4	4	4	4	5	3	3
झ	4	3	4	3	4	5	4	4	3	2	4	4	4	3	4	5	3	4	3	4
ञ	3	2	3	2	3	4	3	3	2	1	3	3	3	2	4	4	4	3	3	3
ट	2	3	2	5	2	3	2	4	3	4	2	2	3	3	4	3	4	4	4	4
ठ	3	4	3	6	3	4	3	5	4	3	3	1	4	2	4	4	3	3	4	5
ड	3	2	3	4	3	2	3	3	3	5	2	3	3	4	4	4	4	5	3	3
ढ	4	3	4	5	4	3	4	4	4	4	3	2	4	3	4	5	3	4	3	4
ण	3	2	3	4	3	2	3	3	3	3	2	1	3	2	4	4	4	3	3	3
~ण	3	3	3	3	3	1	3	2	3	4	3	2	3	3	4	3	4	2	3	2
त	1	2	1	4	1	4	2	3	3	3	3	3	2	2	4	4	4	5	4	5
थ	2	3	2	5	2	5	3	4	4	2	4	2	3	1	4	5	3	4	4	6
द	2	1	2	3	2	3	3	2	3	4	3	4	2	3	4	5	4	6	3	4
ध	3	2	3	4	3	4	4	3	4	3	4	3	3	2	4	6	3	5	3	5
न	2	1	2	3	2	3	3	2	3	2	3	2	2	1	4	5	4	4	3	4
प	2	3	2	3	2	3	1	2	3	4	3	4	3	3	4	3	4	4	4	4

Table 4.8: Etic and Hamming Distances of Consonants-4

	ल		ळ		ळ		व		श		ष		स		क		ख		ग	
फ	3	4	3	4	3	4	2	3	4	3	4	3	4	2	4	4	3	3	4	5
ब	3	2	3	2	3	2	2	1	3	5	3	5	3	4	4	4	4	5	3	3
भ	4	3	4	3	4	3	3	2	4	4	4	4	4	3	4	5	3	4	3	4
म	3	2	3	2	3	2	2	1	3	3	3	3	3	2	4	4	4	3	3	3
य	1	1	1	1	2	3	1	2	2	2	3	4	3	3	4	3	4	4	4	2
र	1	1	1	3	2	1	1	2	3	4	2	2	3	3	4	3	4	4	4	2
ल	0	0	0	2	1	2	1	1	3	3	3	3	2	2	4	4	4	5	4	3
ळ	0	2	0	0	1	2	1	1	3	3	3	5	2	4	4	2	4	3	4	1
ळ	1	2	1	2	0	0	2	1	3	5	3	3	2	4	3	2	3	3	3	1
व	1	1	1	1	2	1	0	0	3	4	3	4	3	3	4	3	4	4	4	2
श	3	3	3	3	3	5	3	4	0	0	1	2	1	1	3	3	3	2	4	4
ष	3	3	3	5	3	3	3	4	1	2	0	0	1	1	3	3	3	2	4	4
स	2	2	2	4	2	4	3	3	1	1	1	1	0	0	3	4	3	3	4	5
क	4	4	4	2	3	2	4	3	3	3	3	3	3	4	0	0	1	1	1	1
ख	4	5	4	3	3	3	4	4	3	2	3	2	3	3	1	1	0	0	2	2
ग	4	3	4	1	3	1	4	2	4	4	4	4	4	5	1	1	2	2	0	0
ज	4	3	4	1	3	3	4	2	3	4	4	6	4	5	2	3	3	4	1	2
झ	4	4	4	2	3	4	4	3	3	3	4	5	4	4	3	4	2	3	2	3
tc	4	4	4	2	3	4	4	3	2	3	3	5	3	4	1	2	2	3	2	3
ड	4	3	4	3	3	1	4	2	4	6	3	4	4	5	2	3	3	4	1	2
र	4	3	4	3	3	1	4	2	4	4	4	2	4	3	2	1	3	2	1	2
ढ	3	4	3	4	2	2	4	3	4	5	4	3	3	4	3	4	2	3	2	3
फ़	4	3	4	3	3	3	3	2	3	2	3	2	3	1	2	3	1	2	3	4

Table 4.9: Etic and Hamming Distances of Consonants-5

	ज	झ	tc	ड	र	ढ	फ़
ह	3	5	2	4	2	4	3
:	2	4	1	3	3	5	2
क	4	2	4	3	4	1	4
ख	4	3	3	2	4	2	4
ग	3	1	3	2	4	2	3
घ	3	2	2	1	4	3	3
ङ	3	3	3	2	4	4	3
च	3	2	3	3	3	1	4
छ	3	3	2	2	3	2	4
ज	2	1	2	2	3	2	3
झ	2	2	1	1	3	3	4
ञ	2	3	2	2	3	4	5
ट	4	4	4	5	4	3	4
ठ	4	5	3	4	4	4	3
ड	3	3	3	4	4	4	5
ढ	3	4	2	3	4	5	4
ण	3	5	3	4	4	6	3

Table 4.9: Etic and Hamming Distances of Consonants-5

	ज		झ		tc		ड़		र		ढ़		फ़	
ण	3	4	3	3	4	5	3	2	2	2	3	1	4	2
त	4	3	4	4	4	2	4	3	4	3	3	4	4	3
थ	4	4	3	3	4	3	4	4	4	4	2	3	3	2
द	3	2	3	3	4	3	3	2	3	4	2	3	4	4
ध	3	3	2	2	4	4	3	3	3	5	1	2	3	3
न	3	4	3	3	4	5	3	4	3	4	2	3	4	2
प	4	2	4	3	4	1	4	2	4	2	4	3	3	2
फ	4	3	3	2	4	2	4	3	4	3	3	2	2	1
ब	3	1	3	2	4	2	3	1	3	3	3	2	3	3
भ	3	2	2	1	4	3	3	2	3	4	2	1	2	2
म	3	3	3	2	4	4	3	3	3	3	3	2	3	1
य	3	2	3	3	3	3	4	4	4	4	4	5	4	4
र	4	4	4	5	4	5	3	2	4	2	4	3	4	4
ल	4	3	4	4	4	4	4	3	4	3	3	4	4	3
ळ	4	1	4	2	4	2	4	3	4	3	3	4	4	3
ळ	3	3	3	4	3	4	3	1	3	1	2	2	3	3
व	4	2	4	3	4	3	4	2	4	2	4	3	3	2
श	3	4	3	3	2	3	4	6	4	4	4	5	3	2
ष	4	6	4	5	3	5	3	4	4	2	4	3	3	2
स	4	5	4	4	3	4	4	5	4	3	3	4	3	1
क	2	3	3	4	1	2	2	3	2	1	3	4	2	3
ख	3	4	2	3	2	3	3	4	3	2	2	3	1	2
ग	1	2	2	3	2	3	1	2	1	2	2	3	3	4
ज	0	0	1	1	1	1	1	2	1	4	2	3	3	4
झ	1	1	0	0	2	2	2	3	2	5	1	2	2	3
tc	1	1	2	2	0	0	2	3	2	3	3	4	2	3
ड़	1	2	2	3	2	3	0	0	1	2	2	1	3	4
र	1	4	2	5	2	3	1	2	0	0	2	3	3	2
ढ़	2	3	1	2	3	4	2	1	2	3	0	0	2	3
फ़	3	4	2	3	2	3	3	4	3	2	2	3	0	0

ZO

Appendix A

Panini's grammar for Sanskrit

The following *slokas* are taken from [8] . I quote from the preface of the book:

The Groundwork of the grammatical literature of the Sanskrit is comprised in Pāṇini's eight Lectures entitled "The Ashtādhyāyī". Each of the lectures is divided into a number of *sūtras*, or succinct aphorisms. On these Mr. Colebrooke remarks :- " The studied brevity of the *Pāṇiniyansūtras* renders them in the highest degree obscure; even with the knowledge of the key to their interpretation, the student finds them ambiguous. In the application of them when understood, he discovers many seeming contradictions; and with every exertion of practised memory, he must experience the utmost difficulty in combining rules dispersed in apparent confusion through different portions of Pāṇini's eight lectures."

Quoted below are Pāṇini's *sūtras* 1, 15 and 16 giving the Pāṇinian alphabet order, classification and origin of Vowels and Consonants respectively:

sūtra 1

अ इ उण् । १ । ऋ लृक् । २ । ए औङ् । ३ । ऐ औच् । ४ । ह व रत् । ५ ।
लण् । ६ । ज म ङ ण नम् । ७ । झ भञ् । ८ । घ ढ धष् । ९ । ज व ग ङ
दश् । १० । ख फ छ ठ थ च ट तव् । ११ । क पय् । १२ । श ष सर् । १३ ।
हल् । १४ ।
इति माहेश्वराणि सूचाण्यणादिसंज्ञार्थानि ।

This sutra gives the Sanskrit alphabet and its arrangement according to Panini. Interestingly, it differs a lot from our modern arrangement. Actually, in Indian aesthetics, there were two distinct arrangements called 'Maalini' and 'Maatrika'. 'Maatrika' is similiar to our modern phonetic arrangement . Pāṇini's arrangement is supposed to be closer to the 'Maalini' system. The basis of the 'Maalini' system is believed to be lost.

sūtra 15

अकुहविसर्जनीयानां कण्ठः । इचुयशानां तालु । ऋटुरषाणां मूर्धा ।
लृटुलसानां दन्ताः । उपपध्मानीयानामोष्ठौ । जमङणनानां नासिका च

। एदैतोः कण्ठतालु । औदौतोः कण्ठोष्ठम् । वकारस्य दन्तोष्ठम् ।
जिह्वामूलीयस्य जिह्वामूलम् । नासिकानुस्वारस्य ।

This sutra tells about the articulation of Vowels. Consonant articulation is described by the next sutra (sutra 16).

Appendix B

Source Code for helper programs

The file `defs.h` header file contains the definitions used for classifying the phones, construction of the etic systems and computation of etic distances. The Program `main.c` is used for calculating the distances and outputting them as L^AT_EX code previously used in this text. The program `charset.c` gives the Figure 3.1 and from them the Figure 3.3 and Figure 3.2.

```

/*****
                                defs.h  -  description
                                -----
begin                          : Wed Nov 20 2002
copyright                      : (C) 2002 by Amitabh Trehan
email                          : amitabhtrehan@softhome.net
*****/

/*****
*
*   This program is free software; you can redistribute it and/or modify
*   it under the terms of the GNU General Public License as published by
*   the Free Software Foundation; either version 2 of the License, or
*   (at your option) any later version.
*
*****/

#define NUMPARTS 6
#define NUMVOWELS 20 // Number of Vowels
#define NUMCONS 47 // Number of Consonants
#define MAXP 50 // Max phones assumed

typedef enum {false, true} bool;

// 0 for any feature means no feature is selected from that partition. This is default in
// Vowel Feature sets
```

```

typedef enum { low =1 , mid, high} VF1;
typedef enum { front =1, central, back } VF2;
typedef enum { unrounded =1, rounded } VF3;
typedef enum {shortv =1, longv, diphthong} VF4;
typedef enum {vibrant=1, lateral} VF5;

//Consonant Feature sets

typedef enum {stop =1, fricative, affricate, nasal, sibilant,liquid=10, glide=10} CF1;
typedef enum {glottal = 1, velar, palatal, retroflex, alveolar ,dental, labial } CF2;
typedef enum {unvoiced =1, voiced} CF3;
typedef enum {unaspirated =1, aspirated} CF4;
//typedef enum {derived=1} CF5;

typedef struct
{
    char name[4];          //devnag name for DEVANAGARI
    char description[41];  //Short description
    // TODO: find meaningful names for feature sets below
    //
    VF1 f1;
    VF2 f2;
    VF3 f3;
    VF4 f4;
    VF4 f5;
}voweltype;

typedef struct
{
    char name[4];          //devnag name for devanagari
    char description[41];  //Short description

    // TODO: find meaningful names for feature sets below
    //
    CF1 f1;
    CF2 f2;
    CF3 f3;
    CF4 f4;
    //    CF5 f5;
} constype;

typedef struct
{
    char name1[4];
    char name2[20];
} mactable;

typedef struct
{

```

```

    char name1[4];
    unsigned short code;
} codetable;

//int totalphones = 0; // Total character set

short unsigned int voweleticd[NUMVOWELS][NUMVOWELS]; //Etic Distances
short unsigned int vowelhammingd[NUMVOWELS][NUMVOWELS]; // Hamming Distances

short unsigned int conseticd[NUMCONS][NUMCONS]; //Etic Distances
short unsigned int conshammingd[NUMCONS][NUMCONS]; // Hamming Distances

bool hamflag=false; // Is hamming distance calculation on?


voweltype vowel[NUMVOWELS] = // All the Vowels
{
{"a","DEVANAGARI LETTER A", low, central, unrounded, shortv,0},
{"A","DEVANAGARI LETTER AA", low, central, unrounded, longv,0},
{"")a","Bengali/Oriya LETTER Rounded A", mid,central,rounded, shortv,0},
{"i","DEVANAGARI LETTER I", high, front, unrounded, shortv,0},
{"I","DEVANAGARI LETTER II", high, front, unrounded, longv,0},
{"u","DEVANAGARI LETTER U", high, back,rounded, shortv,0},
{"U","DEVANAGARI LETTER UU", high, back,rounded, longv,0},
{".r","DEVANAGARI LETTER VOCALIC R", high, front,unrounded,shortv,vibrant},
{".R","DEVANAGARI LETTER VOCALIC RR", high, front,unrounded,longv,vibrant},
{".ru","Oriya LETTER VOCALIC R", high, back,rounded,shortv,vibrant},
{".l","DEVANAGARI LETTER VOCALIC L", high, front,unrounded,shortv,lateral},
{".L","DEVANAGARI LETTER VOCALIC LL", high, front,unrounded,longv,lateral},
{"")e","DRAVIDIAN LETTER CANDRA E", mid,front,rounded,shortv,0},
{"<e","DRAVIDIAN LETTER SHORT E", mid,front,unrounded,shortv,0},
{"e","DEVANAGARI LETTER E", mid,front,unrounded,longv,0},
{"ai","DEVANAGARI LETTER AI", 0,front,unrounded,diphthong,0},
{"")o","DEVANAGARI LETTER CANDRA O", low, central, rounded, shortv,0},
{"<o","DRAVIDIAN LETTER SHORT O", mid,back,rounded,shortv,0},
{"o","DEVANAGARI LETTER O", mid,back,rounded,longv,0},
{"au","DEVANAGARI LETTER AU", 0,back,rounded,diphthong,0},
};

constype consonant[NUMCONS]= // All the Consonants
{
{"ha","DEVANAGARI LETTER HA",fricative, glottal,unvoiced,aspirated},
{"h","DEV LETTER H (Visarga)",fricative, glottal,voiced,aspirated},
{"ka","DEVANAGARI LETTER KA",stop, velar,0,0},
{"kha","DEVANAGARI LETTER KHA",stop, velar,0,aspirated},
{"ga","DEVANAGARI LETTER GA",stop, velar,voiced,0},

```

```

{"gha","DEVANAGARI LETTER GH A",stop, velar,voiced,aspirated},
{"nga","DEVANAGARI LETTER NG A",nasal, velar,voiced,0},
{"ca","DEVANAGARI LETTER CA",affricate, palatal,0,0},
{"cha","DEVANAGARI LETTER CH A",affricate, palatal,0,aspirated},
{"ja","DEVANAGARI LETTER JA",affricate, palatal,voiced,0},
{"jha","DEVANAGARI LETTER JH A",affricate, palatal,voiced,aspirated},
{"nya","DEVANAGARI LETTER NY A",nasal, palatal,voiced,0},
{".ta","DEVANAGARI LETTER TT A",stop, retroflex,0,0},
{".tha","DEVANAGARI LETTER TT HA",stop, retroflex,0,aspirated},
{".da","DEVANAGARI LETTER DD A",stop, retroflex,voiced,0},
{".dha","DEVANAGARI LETTER DD HA",stop, retroflex,voiced,aspirated},
{"nna","DEVANAGARI LETTER NN A",nasal, retroflex,voiced,0},
{"nnna","DEVANAGARI (Tamil) LETTER NNNA",nasal, alveolar,voiced,0},
{"ta","DEVANAGARI LETTER TA",stop, dental,0,0},
{"tha","DEVANAGARI LETTER TH A",stop, dental,0,aspirated},
{"da","DEVANAGARI LETTER DA",stop, dental,voiced,0},
{"dha","DEVANAGARI LETTER DH A",stop, dental,voiced,aspirated},
{"na","DEVANAGARI LETTER NA",nasal, dental,voiced,0},
{"pa","DEVANAGARI LETTER PA",stop, labial,0,0},
{"pha","DEVANAGARI LETTER PH A",stop, labial,0,aspirated},
{"ba","DEVANAGARI LETTER BA",stop, labial,voiced,0},
{"bha","DEVANAGARI LETTER BH A",stop, labial,voiced,aspirated},
{"ma","DEVANAGARI LETTER MA",nasal, labial,voiced,0},
{"ya","DEVANAGARI LETTER YA",glide, palatal,0,0},
{"ra","DEVANAGARI LETTER RA",liquid, retroflex,0,0},
{"la","DEVANAGARI LETTER LA",liquid, dental,0,0},
{"llla","DEVANAGARI(Tamil) LLLA",glide, dental,0,0},
{"La","DEVANAGARI(Marathi) LETTER LLA",fricative, dental,0,0},
{"va","DEVANAGARI LETTER VA",glide, labial,0,0},
{"sha","DEVANAGARI LETTER SHA",sibilant, palatal,unvoiced,0},
{"Sha","DEVANAGARI LETTER SSA",sibilant, retroflex,unvoiced,0},
{"sa","DEVANAGARI LETTER SA",sibilant, dental,unvoiced,0},
{"qa","DEVANAGARI LETTER QA",fricative, velar,unvoiced,unaspirated},
{"Kha","DEVANAGARI LETTER KHHA",fricative, velar,unvoiced,aspirated },
{"Ga","DEVANAGARI LETTER GHHA",fricative, velar,voiced,unaspirated },
{"za","DEVANAGARI LETTER ZA",fricative, palatal,voiced,unaspirated },
{"Jha","DRAVIDIAN LETTER JHA",fricative, palatal,voiced,aspirated },
{"tc","DRAVIDIAN(Telegu) LETTER TC",fricative, palatal,unvoiced,unaspirated },
{"Ra","DEVANAGARI LETTER DDDHA",fricative, retroflex,voiced,unaspirated },
{"RRa","DEVANAGARI(Tamil) LETTER RRA",fricative, alveolar,voiced,unaspirated },
{"Rha","DEVANAGARI LETTER RHA",fricative, dental,voiced,aspirated },
{"fa","DEVANAGARI LETTER FA",fricative, labial,unvoiced,aspirated },
};

```

```

// Vowel Print Table
maptable vTeXtab[ NUMVOWELS ] =
{
    {"a","{\n a}"},
    {"A","{\n A}"},
    {"")a","\\textipa{0}"},

```

```

{"i","{\n i}"},
{"I","{\n I}"},
{"u","{\n u}"},
{"U","{\n U}"},
{"r","{\n .r}"},
{"R","{\n .R}"},
{"ru","ru"},
{"l","{\n .l}"},
{"L","{\n .L}"},
{"e","{\n e}"},
{"<e","{\textipa{e; } }"},
{"e","{\n e}"},
{"ai","{\n ai}"},
{"o","{\n ~o}"},
{"<o","{\textipa{o; } }"},
{"o","{\n o}"},
{"au","{\n au}"},

};
maptable cTeXtab[NUMCONS]=
{
    {"ha","{\n ha}"},
    {"h","{\n .h}"},
    {"ka","{\n ka}"},
    {"kha","{\n kha}"},
    {"ga","{\n ga}"},
    {"gha","{\n gha}"},
    {"nga","{\n \na}"},
    {"ca","{\n ca}"},
    {"cha","{\n cha}"},
    {"ja","{\n ja}"},
    {"jha","{\n jha}"},
    {"nya","{\n ~na}"},
    {"ta","{\n .ta}"},
    {"tha","{\n .tha}"},
    {"da","{\n .da}"},
    {"dha","{\n .dha}"},
    {"nna","{\n .na}"},
    {"nnna","{\n ~{\n .na}}"},
    {"ta","{\n ta}"},
    {"tha","{\n tha}"},
    {"da","{\n da}"},
    {"dha","{\n dha}"},
    {"na","{\n na}"},
    {"pa","{\n pa}"},
    {"pha","{\n pha}"},
    {"ba","{\n ba}"},
    {"bha","{\n bha}"},
    {"ma","{\n ma}"},
    {"ya","{\n ya}"},

```

```

{"ra","{\n ra}"},
{"la","{\n la}"},
{"lla","{\d{\n La}}"},
{"La","{\n La}"},
{"va","{\n va}"},
{"sha","{\n \"sa\"},},
{"Sha","{\n .sa}"},
{"sa","{\n sa}"},
{"qa","{\n qa}"},
{"Kha","{\n .kha}"},
{"Ga","{\n .ga}"},
{"za","{\n za}"},
{"Jha","{\d{\n jha}}"},
{"tc","{tc}"},
{"Ra","{\n Ra}"},
{"RRa","{\d{\n ra}}"},
{"Rha","{\n Rha}"},
{"fa","{\n fa}"},
};

```

```

codetable vcodetab[NUMVOWELS]=

```

```

{
    {"a",0x04},
    {"A",0x05},
    {"")a",0x06},
    {"i",0x08},
    {"I",0x09},
    {"u",0x0e},
    {"U",0x0f},
    {"".r",0x14},
    {"".R",0x15},
    {"".ru",0x12},
    {"".l",0x16},
    {"".L",0x17},
    {"")e",0x1a},
    {""<e",0x18},
    {""e",0x19},
    {""ai",0x1b},
    {"")o",0x07},
    {""<o",0x1c},
    {""o",0x1d},
    {""au",0x1f}

```

```

};

```

```

codetable ccodetab[NUMCONS]=

```

```

{
    {"ha",0x45},
    {".h",0x47},
    {"ka",0x48},
    {"kha",0x49},

```

```

{"ga",0x4a},
{"gha",0x4b},
{"nga",0x4f},
{"ca",0x50},
{"cha",0x51},
{"ja",0x52},
{"jha",0x53},
{"nya",0x57},
{"ta",0x60},
{"tha",0x61},
{"da",0x62},
{"dha",0x63},
{"nna",0x67},
{"nnna",0x6f},
{"ta",0x70},
{"tha",0x71},
{"da",0x72},
{"dha",0x73},
{"na",0x77},
{"pa",0x78},
{"pha",0x79},
{"ba",0x7a},
{"bha",0x7b},
{"ma",0x7f},
{"ya",0x56},
{"ra",0x66},
{"la",0x76},
{"llla",0x5e},
{"La",0x6e},
{"va",0x7e},
{"sha",0x55},
{"Sha",0x65},
{"sa",0x75},
{"qa",0x4c},
{"Kha",0x4d},
{"Ga",0x4e},
{"za",0x5a},
{"Jha",0x5b},
{"tc",0x58},
{"Ra",0x6a},
{"RRa",0x6c},
{"Rha",0x6b},
{"fa",0x7d}
};

```

```

/*****
                                main.c - description
                                -----
begin                          : Thu Nov 14 00:03:01 IST 2002
copyright                      : (C) 2002 by Amitabh Trehan
email                          : amitabhtrehan@softhome.net
*****/

/*****
*
*   This program is free software; you can redistribute it and/or modify
*   it under the terms of the GNU General Public License as published by
*   the Free Software Foundation; either version 2 of the License, or
*   (at your option) any later version.
*
*****/

#ifdef HAVE_CONFIG_H
#include <config.h>
#endif

#include <stdio.h>
#include <stdlib.h>
#include "defs.h"

#define NCOLS 10                // Vertical division of consonant tables

FILE *outfptr;

int main(int argc, char *argv[])
{

    int ctr, x,y,z ;

    //Initialisation

    for(x=0; x<NUMVOWELS; x++)
    {
        for(y=0; y< NUMVOWELS; y++)
        {
            voweleticd[x][y] = 0;
            vowelhammingd[x][y] = 0;
        }
    }

    for(x=0; x<NUMCONS; x++)
    {
        for(y=0; y< NUMCONS; y++)

```

```

        {
            conseticd[x][y] = 0;
            conshammingd[x][y] = 0;
        }
    }

// Processing for Vowels

//Calculate the etic distances

for(x=0; x< NUMVOWELS ; x++)
{
    for(y=x+1; y< NUMVOWELS; y++)
    {
        ctr = 0;
        if(vowel[x].f1 != vowel[y].f1) ctr++;
        if(vowel[x].f2 != vowel[y].f2) ctr++;
        if(vowel[x].f3 != vowel[y].f3) ctr++;
        if(vowel[x].f4 != vowel[y].f4) ctr++;
        if(vowel[x].f5 != vowel[y].f5) ctr++;

        // Transfer the value
        voweleticd[x][y] = voweleticd[y][x] = ctr;
        // Compute hamming distance

        vowelhammingd[x][y] = vowelhammingd[y][x] = hamd(vcodetab[x].code,vcodetab[y].code);
    }
}

//Processing for consonants

for(x=0; x< NUMCONS ; x++)
{
    for(y=x+1; y< NUMCONS; y++)
    {
        ctr = 0;
        if(consonant[x].f1 != consonant[y].f1) ctr++;
        if(consonant[x].f2 != consonant[y].f2) ctr++;
        if(consonant[x].f3 != consonant[y].f3) ctr++;
        if(consonant[x].f4 != consonant[y].f4) ctr++;
        // if(consonant[x].f5 != consonant[y].f5) ctr++;

        // Transfer the value
        conseticd[x][y] = conseticd[y][x] = ctr;
        // Compute hamming distance
        conshammingd[x][y] = conshammingd[y][x] = hamd(ccodetab[x].code,ccodetab[y].code);
    }
}

```

```

    printvdists();
    printcdists();

    //Write into TeX file

    //Open Table Output file which is a TeX file

/*    if((outfptr = fopen("etictable.dn","w")) == NULL)
    {
        printf("Cannot open output file for writing \n");
        exit(-1);
    }
*/

    writeTeX();
    writehamTeX();

    // Close files

}

int writeTeX()
{
    int ctr,x,y, ncols, ycount;
    //Open Table Output file which is a TeX file

    if((outfptr = fopen("etictable.dn","w")) == NULL)
    {
        printf("Cannot open output file for writing \n");
        exit(-1);
    }

    fprintf(outfptr, "\\documentclass{article} \n");
    fprintf(outfptr, "\\usepackage{dev} \n");
    fprintf(outfptr, "\\usepackage{tipa} \n");
    fprintf(outfptr, "\\usepackage{longtable} \n");
    //    fprintf(outfptr, "@tabs \n");
    fprintf(outfptr, "\\begin{document}\n { \n");

    //Write Vowel Table

    fprintf(outfptr, "\\begin{longtable}{|");
    for(ctr=0;ctr<=NUMVOWELS;ctr++)
    {

```



```

        fprintf(outfptr, "l|");
    }
    ctr--;    // Set counter correctly
    fprintf(outfptr, "}\n");
    fprintf(outfptr, "\\caption{Etic Distances of Consonants-%d \\label{C%deticd}}\\

    // An empty slot
    fprintf(outfptr,"{ } & ");
    for(;x< ncols-1;x++)
    {
        fprintf(outfptr,"%s& ",cTeXtab[x].name2);
    }
    // Last element
    fprintf(outfptr,"%s \\\\ \\hline \\hline ",cTeXtab[x].name2);
    fprintf(outfptr,"\n");
    fprintf(outfptr," \\endhead \\");
    fprintf(outfptr,"\n");

    for(x=0;x< NUMCONS;x++)
    {
        fprintf(outfptr,"%s& ",cTeXtab[x].name2);
        for(y=ycount;y< ncols-1;y++)
        {
            fprintf(outfptr,"%d & ",conseticd[x][y]);
        }
        // Last element
        fprintf(outfptr,"%d ",conseticd[x][y]);
        fprintf(outfptr,"\\\\ \\hline");
        fprintf(outfptr,"\n");
    }
    fprintf(outfptr,"\\hline");
    fprintf(outfptr,"\n");
    fprintf(outfptr, "\\end{longtable}");
    fprintf(outfptr,"\n");

    if(ncols == NUMCONS) break;
    ncols = (ncols+NCOLS) > NUMCONS ? NUMCONS : ncols+10 ;
}

// Close devnag TeX Document

fprintf(outfptr, "\\end{document}");

fclose(outfptr);

}

int writehamTeX()

```

```

{
    int ctr,x,y, ncols, ycount;
    FILE *hfptr;

    //Open Table Output file which is a TeX file

    if((hfptr = fopen("hamtable.dn","w")) == NULL)
    {
        printf("Cannot open output file for writing \n");
        exit(-1);
    }

    //Write Vowel Tables

    fprintf(hfptr, "\\begin{longtable}{|l|}");
    for(ctr=0;ctr<=10;ctr++)
    {
        fprintf(hfptr, "l|l|");
    }
    fprintf(hfptr, "}\n");
    fprintf(hfptr, "\\caption{Etic and Hamming Distances of Vowels-I \\label{vowelet

        // An empty slot
    fprintf(hfptr," & ");

    for(x=0;x<=10;x++)
    {
        fprintf(hfptr,"\\multicolumn{2}{c|}{%s}& ",vTeXtab[x].name2);
    }
    // Last element
    fprintf(hfptr,"\\multicolumn{2}{c|}{%s} \\\\ \\hline \\hline ",vTeXtab[x].name2);
    fprintf(hfptr," \\\\ \\hline \\hline ");
    fprintf(hfptr,"\\n");
    fprintf(hfptr," \\endhead \\n");

    for(x=0;x< NUMVOWELS;x++)
    {
        fprintf(hfptr,"%s& ",vTeXtab[x].name2);
        for(y=0;y<10;y++)
        {
            fprintf(hfptr,"%d & ",voweleticd[x][y]);
            fprintf(hfptr,"%d & ",vowelhammingd[x][y]);
        }
        // Last element
        fprintf(hfptr,"%d & ",voweleticd[x][y]);
        fprintf(hfptr,"%d",vowelhammingd[x][y]);
        fprintf(hfptr,"\\\\ \\hline");
        fprintf(hfptr,"\\n");
    }
}

```

```

    }
    fprintf(hfp, "\\hline");
    fprintf(hfp, "\\end{longtable}");
    fprintf(hfp, "\\n");

fprintf(hfp, "\\begin{longtable}{|l|l|");
    for(;ctr<=NUMVOWELS;ctr++)
    {
        fprintf(hfp, "l|l|");
    }
    fprintf(hfp, "}\\n");
    fprintf(hfp, "\\caption{Etic and Hamming Distances of Vowels-II \\label{vowels}}");

    // An empty slot
    fprintf(hfp, " & ");
    for(x=10;x< NUMVOWELS-1;x++)
    {
        fprintf(hfp, "\\multicolumn{2}{c|}{%s}& ",vTeXtab[x].name2);
    }
    // Last element
    fprintf(hfp, "\\multicolumn{2}{c|}{%s} \\ \\hline \\hline ",vTeXtab[x].name2);
    //fprintf(hfp, " \\ \\hline \\hline ");
    fprintf(hfp, "\\n");
    fprintf(hfp, "\\endhead \\n");

    for(x=0;x< NUMVOWELS;x++)
    {
        fprintf(hfp, "%s& ",vTeXtab[x].name2);
        for(y=10;y< NUMVOWELS-1;y++)
        {
            fprintf(hfp, "%d & ",voweleticd[x][y]);
            fprintf(hfp, "%d & ",vowelhammingd[x][y]);
        }
        // Last element
        fprintf(hfp, "%d & ",voweleticd[x][y]);
        fprintf(hfp, "%d",vowelhammingd[x][y]);
        fprintf(hfp, " \\ \\hline");
        fprintf(hfp, "\\n");
    }
    fprintf(hfp, "\\hline");
    fprintf(hfp, "\\end{longtable}");
    fprintf(hfp, "\\n");

    //Write Consonant Tables

    // Table1 (Consonants 1-10)

    //Calculate columns

```

```

        ncols=NCOLS;    // At least 10 consonants
        ctr=0;          // Set counter for first run
while(1)
{
//      fprintf(hfp, "\n \\pagebreak \n");
    fprintf(hfp, "\\begin{longtable}{|");
    x=ycount=ctr;
    for(;ctr<=ncols;ctr++)
    {
        fprintf(hfp, "l|l|");
    }
    ctr--;    // Set counter correctly
    fprintf(hfp, "}\n");
    fprintf(hfp, "\\caption{Etic and Hamming Distances of Consonants-%d \\label{C%

        // An empty slot
        fprintf(hfp, "{ } & ");
        for(;x< ncols-1;x++)
        {
            fprintf(hfp, "\\multicolumn{2}{c|}{%s}& ", cTeXtab[x].name2);
        }
        // Last element
        fprintf(hfp, "\\multicolumn{2}{c|}{%s} \\ \\ \\ \\hline \\hline ", cTeXtab[x].name2);
        fprintf(hfp, "\n");
        fprintf(hfp, " \\endhead \n");

        for(x=0;x< NUMCONS;x++)
        {
            fprintf(hfp, "%s& ", cTeXtab[x].name2);
            for(y=ycount;y< ncols-1;y++)
            {
                fprintf(hfp, "%d & ", conseticd[x][y]);
                fprintf(hfp, "\\textit{%d} & ", conshammingd[x][y]);
            }
            // Last element
            fprintf(hfp, "%d & ", conseticd[x][y]);
            fprintf(hfp, "\\textit{%d} ", conshammingd[x][y]);
            fprintf(hfp, "\\ \\ \\ \\hline");
            fprintf(hfp, "\n");
        }
        fprintf(hfp, "\\hline");
        fprintf(hfp, "\n");
        fprintf(hfp, "\\end{longtable}");
        fprintf(hfp, "\n");

        if(ncols == NUMCONS) break;
        ncols = (ncols+NCOLS) > NUMCONS ? NUMCONS : ncols+10 ;
    }

```

```

        // Close file
        //
        fclose(hfptr);

}

int printvdists()
{
    int x,y;
    printf("\n Vowel Distances\n");
    printf("    ");
    for(x=0;x< NUMVOWELS;x++)
    {
        printf("%3s ",vowel[x].name);

    }
    printf("\n");
    for(x=0;x< NUMVOWELS;x++)
    {
        printf("%3s ",vowel[x].name);
        for(y=0;y< NUMVOWELS;y++)
        {
            printf("%d   ",voweleticd[x][y]);
            printf("%d   ",vowelhammingd[x][y]);
        }
        printf("\n");
    }
}

}

int printcdists()
{
    int x,y;
    printf("\n Consonant Distances\n");
    printf("    ");
    for(x=0;x< NUMCONS;x++)
    {
        printf("%3s ",consonant[x].name);

    }
    printf("\n");

    for(x=0;x< NUMCONS;x++)
    {
        printf("%3s ",consonant[x].name);

```

```

        for(y=0;y< NUMCONS;y++)
        {
            printf("%d  ",conseticd[x][y]);
        }
        printf("\n");
    }

}

unsigned short int hamd(short unsigned int a, short unsigned int b)
{
    // Hamming distance for first 8 bits

    short unsigned int r, ctr, d;
    r = a ^ b; //compute exclusive-OR
    d=0;
    for(ctr=0;ctr<8;ctr++)
    {
        if(r%2 != 0) d++;
        r >>= 1;    // Shift right by 1 bit (div. by 2)
    }
    return d;
}

```

```

/*****
                                charset.c - description
                                -----
begin                          : Sat Dec 14 20:30:01 IST 2002
copyright                      : (C) 2002 by Amitabh Trehan
email                          : amitabhtrehan@softhome.net
*****/

/*****
*
*   This program is free software; you can redistribute it and/or modify
*   it under the terms of the GNU General Public License as published by
*   the Free Software Foundation; either version 2 of the License, or
*   (at your option) any later version.
*
*****/

#include <stdio.h>
#include <stdlib.h>
#include "defs.h"

#define NCOLS 10                // Number of Chars in a row

FILE *outfptr;

int main(int argc, char *argv[])
{

    int ctr, x,y,z ;

    if( (outfptr = fopen( "charset.dn","w")) == NULL)
    {
        printf("cannot open for writing \n");
        exit(0);
    }

    fprintf(outfptr,"\\[ \n S = \\left\\{ \n \\begin{array}{l} \\[ \n");
    for(ctr=0;ctr<NUMVOWELS;ctr++)
    {
        if((ctr+1)%NCOLS == 0)
        {
            fprintf(outfptr,"%s} \\[ \n",vTeXtab[ctr].name2);
        }
        else
        {if((ctr+1)%NCOLS == 1)
            fprintf(outfptr,"\\mbox{");

```

```

        fprintf(outfptr,"%s, ",vTeXtab[ctr].name2);
    }

}

    if((ctr)%NCOLS != 0)
    {
        fprintf(outfptr,"\b\b } \\\ \n");
    }

for(ctr=0;ctr<NUMCONS;ctr++)
{
    if((ctr+1)%NCOLS == 0)
    {
        fprintf(outfptr,"%s } \\\ \n",cTeXtab[ctr].name2);
    }
    else
    { if((ctr+1)%NCOLS == 1)
        fprintf(outfptr,"\\mbox{");
        fprintf(outfptr,"%s, ", cTeXtab[ctr].name2);
    }

}

    if((ctr)%NCOLS != 0)
    {
        fprintf(outfptr," } \\\ \n");
//        fprintf(outfptr,"\b\b } \\\ \n");
    }

    fprintf(outfptr,"\\end{array}\n");
    fprintf(outfptr,"\\right\\}\n\\]\n");

fclose(outfptr);
}

```

Appendix C

The International Phonetic Alphabet

I.P.A. stands for International Phonetic Association. It tries to standardise the representation and classification of sounds from all the World's languages.

Figure C.1: IPA chart courtesy: [5].

Appendix D

System and development platform Details

Hardware AMD-Duron 833 MHZ, 128 MB RAM.

O.S. RedHat Linux 7.3

Integrated development Environment Kdevelop 2.1 (support for cvs, html documentation etc.)

Editor Gvim ver 6.1

Report Typesetting L^AT_EX (T_EXLive7 installation) with packages devnag (for devanagari support), tipa for special diacritics.

Bibliography

- [[1]] Adrian Akmajian, Richard A. Demers, Ann K. Farmer, and Robert M. Harnish. *Linguistics: An Introduction to Language and Communication*. Prentice hall of india Private Limited, New Delhi, 2001.
- [[2]] W. Sidney Allen. *Phonetics in Ancient India*. Oxford University Press, New York.
- [[3]] http://murray.newcastle.edu.au/users/staff/speech/home_pages/graphics/vclsys2.gif.
- [[4]] <http://www.phon.ox.ac.uk/~jcoleman/organs.jpeg>.
- [[5]] www.arts.gla.ac.uk/ipa/fullchart.html.
- [[6]] <http://cls.coe.utk.edu/lpm/esltoolkit/speechorgans.gif>.
- [[7]] Tista Bagchi. Bengali writing. In Bernard [[10]], pages 399–403.
- [[8]] James R. Ballantyne. *Laghukaumudī of Varadarāja*. MOTILAL BANARSIDASS, Delhi, 1891.
- [[9]] Comrie Bernard, editor. *The World's Major Languages*, chapter Part VI: South Asian Writing Systems, pages 371–429. In Bernard [[10]], 1987.
- [[10]] Comrie Bernard, editor. *The World's Major Languages*. Oxford University Press, 1987.
- [[11]] Barron Brainerd. *Introduction to the Mathematics of Language Study*. American Elsevier Publishing Company, Inc., New York, 1971.
- [[12]] William Bright. The devanagari script. In Bernard [[10]], pages 384 – 390.
- [[13]] William Bright. Kannada and telegu writing. In Bernard [[10]], pages 413–419.
- [[14]] Harjeet Singh Gill. The gurmukhi script. In Bernard [[10]], pages 395–398.
- [[15]] Ian Maddieson. *Patterns of sounds*, chapter Appendix B: Phoneme Charts: Hindi-Urdu, page 270. In [[16]], 1984.
- [[16]] Ian Maddieson. *Patterns of sounds*. Cambridge University Press, New York, 1984.

- [[17]] B.P. Mahapatra. Oriya writing. In Bernard [[10]], pages 404–408.
- [[18]] P.J. Mistry. Gujarati writing. In Bernard [[10]], pages 391–394.
- [[19]] K.P. Mohanan. Malayalam writing. In Bernard [[10]], pages 420–425.
- [[20]] Charles W. Morris. *Signs Language and Behaviour*. Prentice-Hall, Englewood Cliffs, 1946.
- [[21]] Richard G. Salomon. Brahmi and kharoshthi. In Bernard [[10]], pages 373 – 383.
- [[22]] Sanford B. Steever. Tamil writing. In Bernard [[10]], pages 426–428.
- [[23]] U.S.S.R. Academy of Sciences V.V. Ivanov, V.N. Toporov. *Sanskrit*. "Nauka" Publishing House, Central Department of Oriental Literature, Moscow, 1968.