Computational Phonology and Pronunciation Lexicons

Anil Kumar Singh

Assistant Professor

KIIT University, Bhubaneshwar

Computational Phonology

- Models of phonology
- Models of variation
- Letter to phoneme conversion
- Transliteration
- Synchronic and diachronic studies of languages
- Historical linguistics
- And many more: **Pronunciation Lexicon**

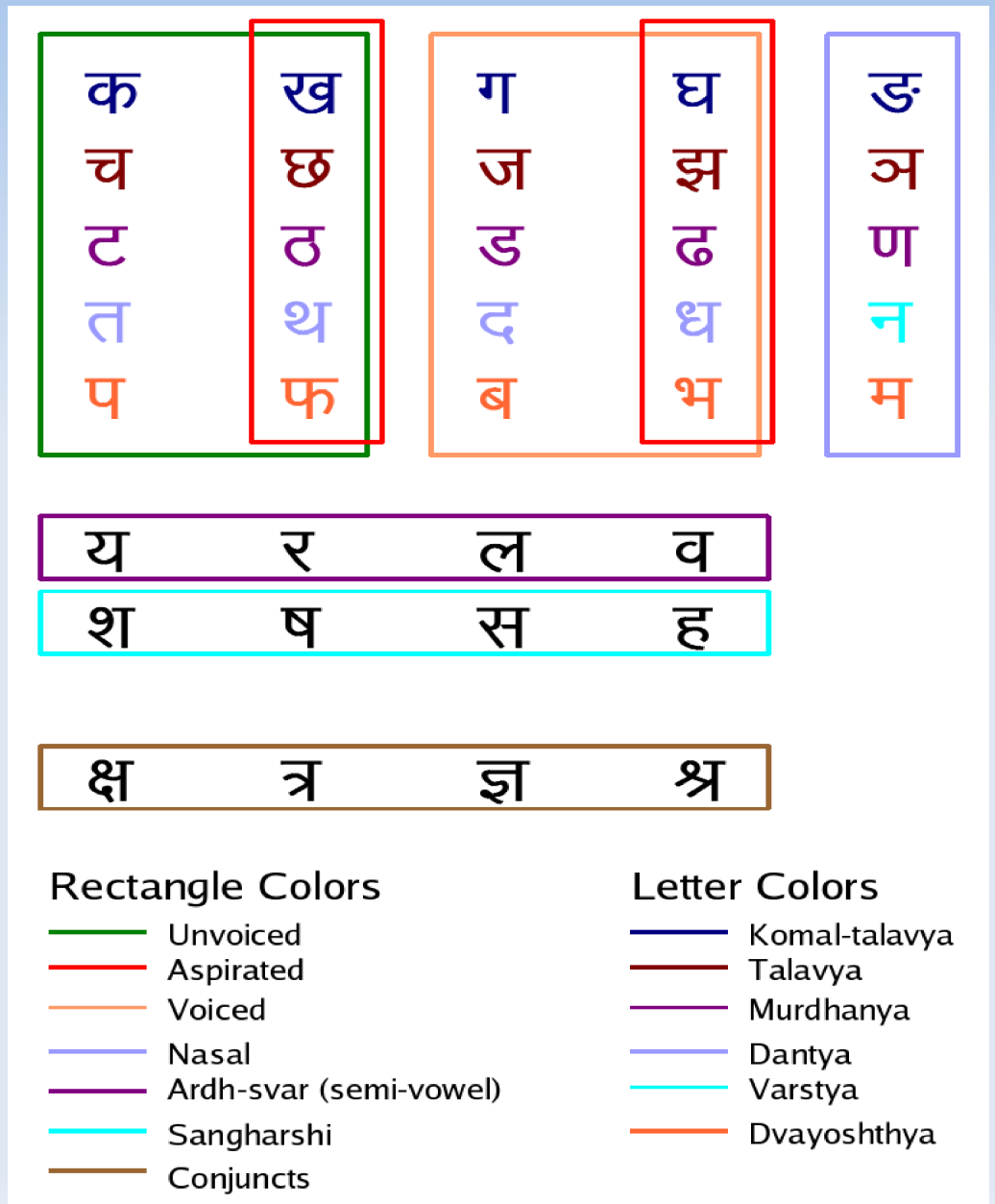
Models of Phonology

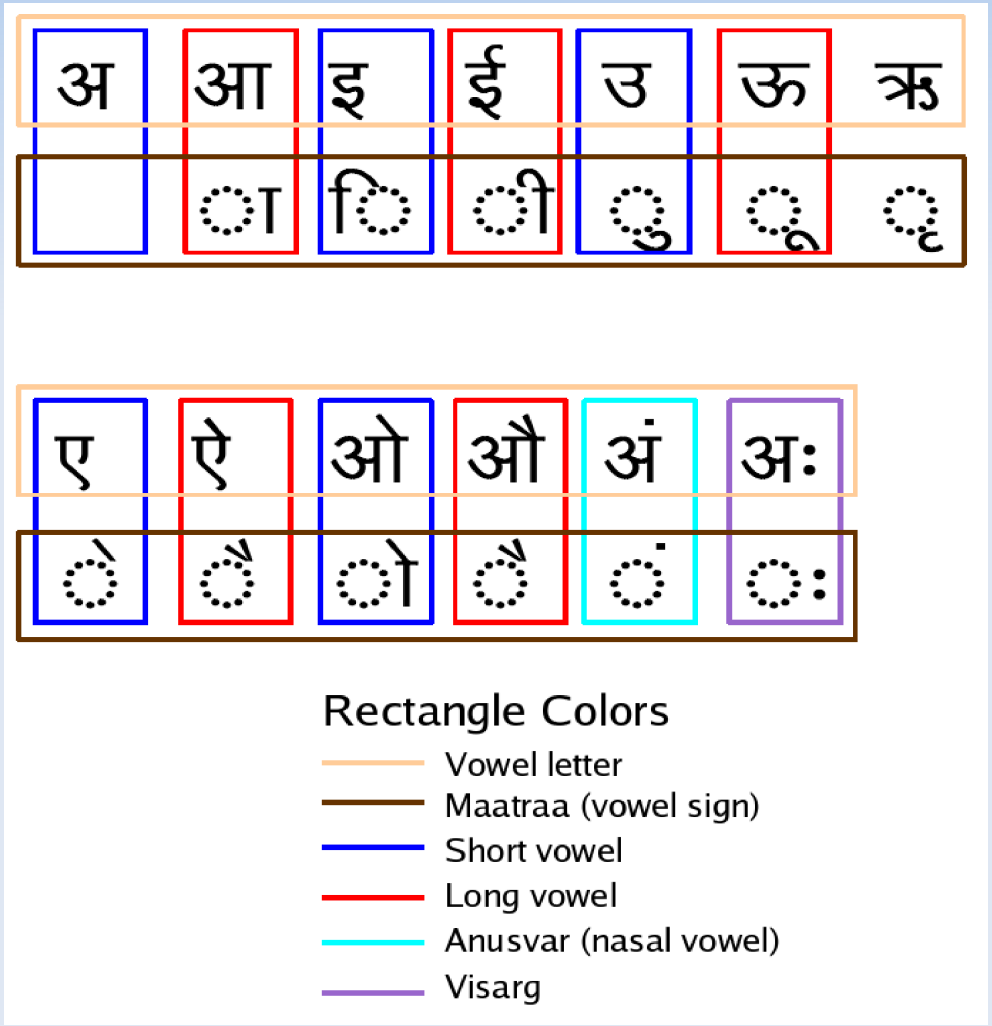
- Articulatory (phonetic) features
- Description of phonemes in terms of features
- Relation of writing systems to phonology

Computational Phonetic Model of Scripts (CPMS)

- Relates phonology and writing systems
- Consists of component models
  - Model of Alphabet
  - Model of Phonology
  - Aaksharik Model
  - Model of Variation
  - Stepped Distance Function
  - An alignment algorithm to estimate 'surface similarity'

Model of Alphabet



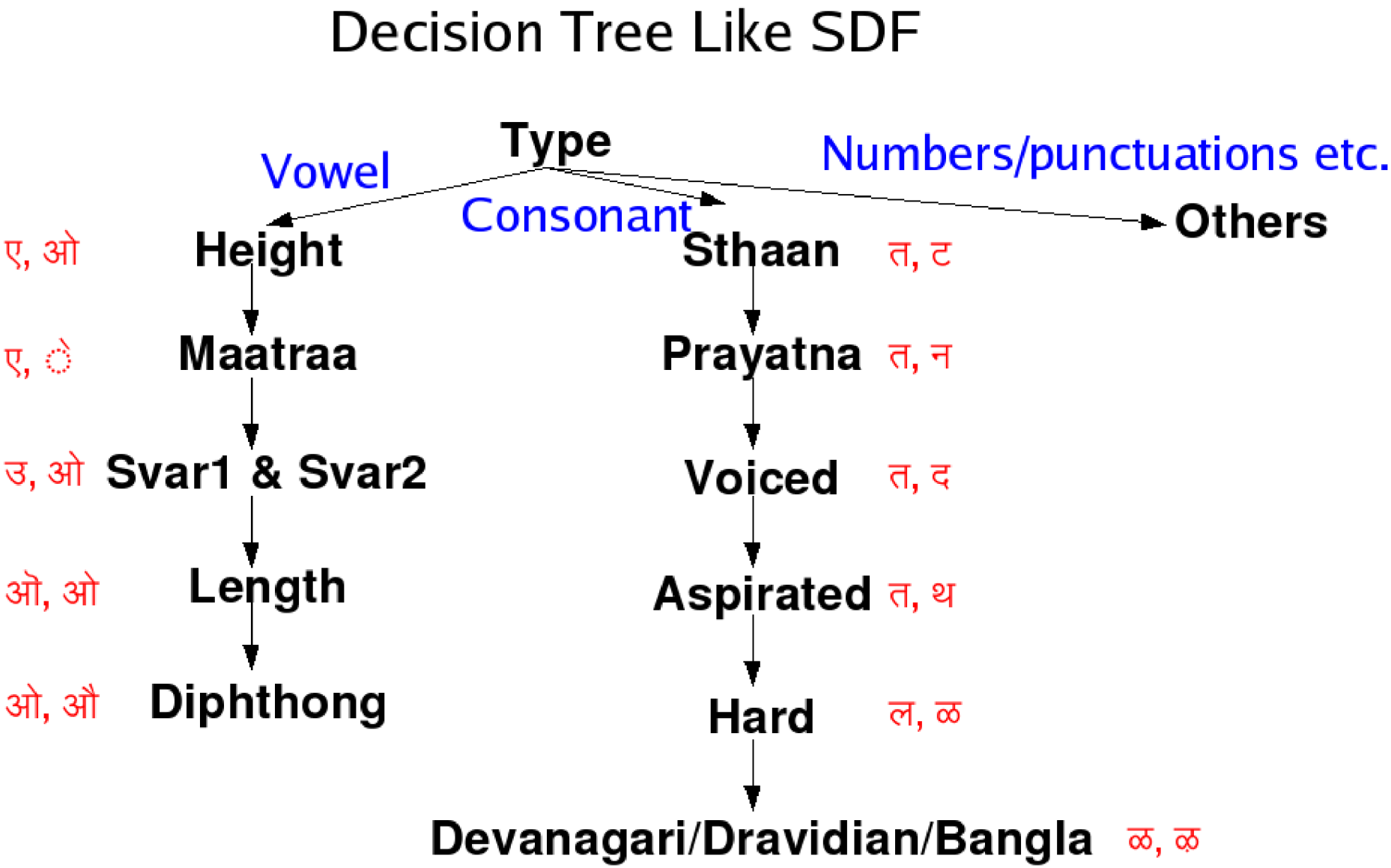


Model of Phonology

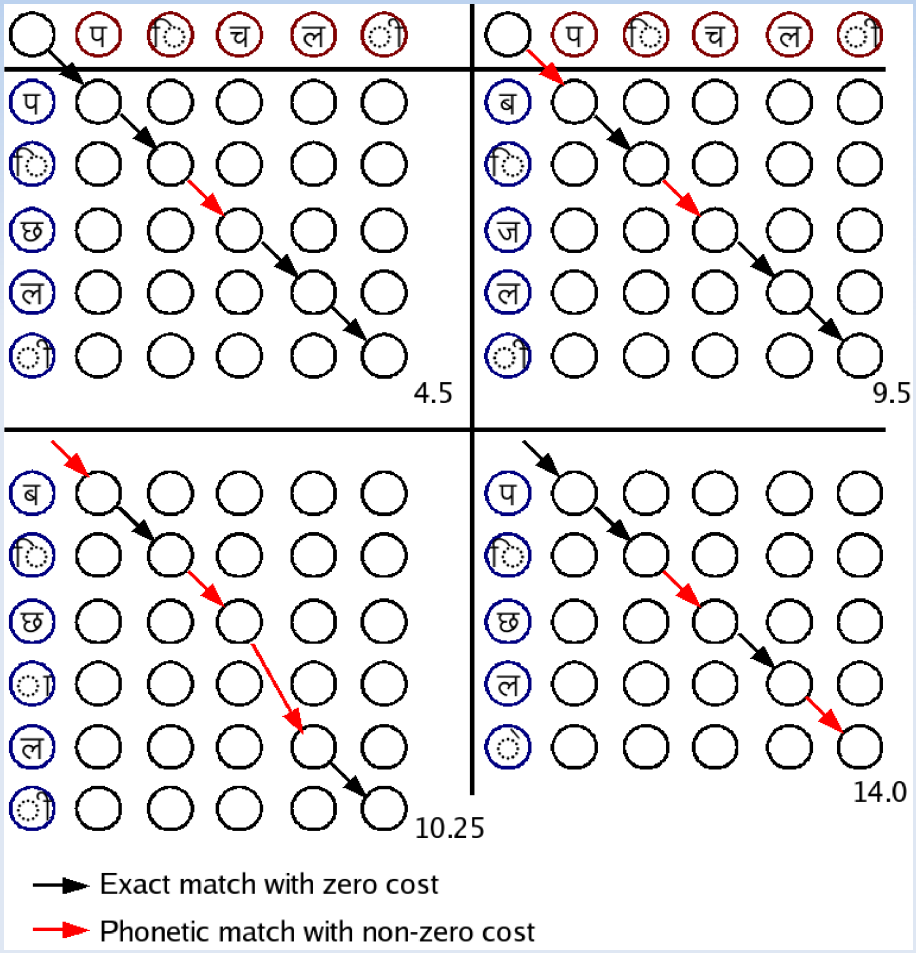
Feature	Possible Values
Type	**C**onsonant, **V**owel, Vowel **m**odifier, Nu**k**ta, **N**umber, **P**unctuation, **H**alant, **U**nused
Height	**F**ront, **M**id, **B**ack
Length	**L**ong, **S**hort, **M**edium
Svar1	**L**ow, Lower Mi**d**dle, Upper, **M**iddle, Lower Hi**g**h, **H**igh
Svar2	**S**amvrit, Ardh-Sa**m**vrit, Ardh-V**i**vrit, **V**ivrit
Place	Dva**y**oshthya (Bilabial), Dant**o**shthya (Labio-dental), **D**antya (Dental), **V**arstya (Alveolar), **T**alavya (Palatal), **M**urdhanya (Retroflex), **K**omal-Talavya (Velar), **J**ivhaa-Muliya (Uvular), **S**varyantramukhi (Pharynxial)
Manner	**S**parsha (Stop), **N**asikya (Nasal), **P**arshvika (Lateral), P**r**akampi (Voiced), Sang**h**arshi (Fricative), Ardh-S**v**ar (Semi-vowel)

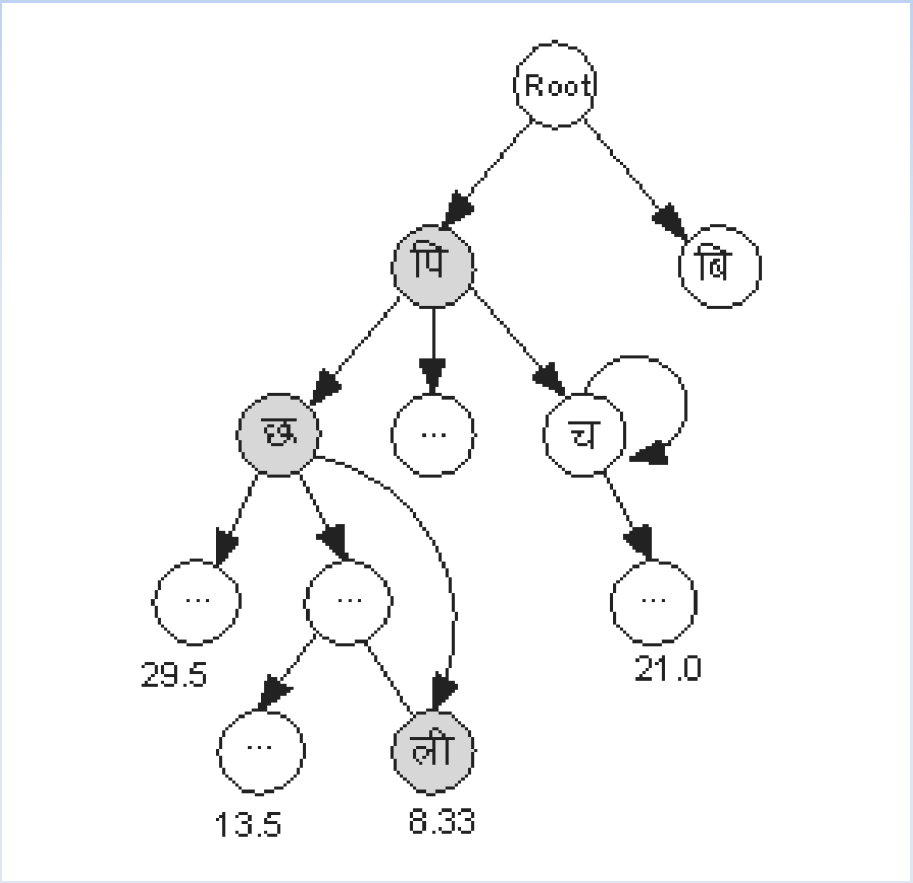
Table 1: Non-Boolean Phonetic Features

Stepped Distance Function



Alignment Algorithm





Models of Variation

- Phonetic variation
- Relation to spelling variation
- Similarities and differences across languages

Letter to Phoneme Conversion

- Given text in a writing system, generate the phonetic representation
- Use of statistical techniques
- Can be modelled as a 'translation' problem

Transliteration

- Conversion from one writing system to another
- In case of Indian languages:
  - Writing systems are highly phonetic in nature
  - Thus, transliteration can use techniques from computational phonology

Synchronic and Diachronic Studies of Languages

- Large amounts of text is available in non-phonetic representation (writing system)
- Converting this text into phonetic representation can be used for:
  - Synchronic and diachronic differences and variation
- E.g. we can calculate linguistic differences among languages
  - Over time (diachronic)
  - And over space (synchronic)

Historical Linguistics

- With data in phonetic representation:
  - Using models of phonology and computational techniques, we can try to derive sound change patterns across languages and dialects
  - Empirical validation of the laws of sound change

Pronunciation Lexicon

- Urgently required for Indian language
- Computational phonology techniques can be used for this purpose
- Basically letter to phoneme conversion

Conclusion

- Standardization can help designing better computational phonology techniques
- Computational phonology can help in building better tools or resources that use the standards