Study of Various Online IPA Convertors Available for Regional Languages and Phonetics of Punjabi Language: A Review

Samandeep Kaur, Student *of M.Tech ,Dept. Of Computer Engineering Punjabi University Patiala*
Samandeep.kaleka@yahoo.com
Charanjiv Singh Saroa, *Assistant Professor, Dept. Of Computer Engineering , Punjabi University* Patiala.
Charanjiv_saroa@yahoo.com

ABSTRACT:
IPA stands for International Phonetic Alphabet. Different languages have different sounds associated with them. IPA is the only way to put sounds in symbols in such a way that any word can be read in the same way anywhere around the globe. It makes spelling clearer and more understandable .In this paper, I have discussed different IPA Convertors available in different languages. In future I will make a Punjabi to IPA convertor.

Keywords:
Phonetics; IPA; phonemes; language transliteration; vowels; consonants.

I.   INTRODUCTION:
Conversion of the text to IPA conversion involves the following steps:

1. Enter the text
2. Process Step no 3 and 4 for each word or character
3. Compare the word/character with database.
4. Replace the word/character with the IPA symbol, provided in the database.

Language transliteration is one of the important areas in natural language processing. Transliteration is the process of mapping a written word from a language-script pair to another language-script pair. It is the computer based conversion of a character or word from one language/script to another without losing its phonological characteristics.

II.   PHONETICS
Phonetics is the science that is used to formulate general laws relating to the sounds and their production. Speech sounds can be examined corresponding to the stages of transmission of the speech signal from a speaker to the listener. Study of sound production by the human organs (like larynx, velum ,tongue, lips etc) comes under the type of phonetics which is known as articulatory phonetics. it may be learnt that there is a difference between phonetics and phonology as phonetics deals with all types of sound production by human beings and phonology is concerned with the sound production of a particular language only. Two types of phonemes are usually used- one is segmental phonemes and the other one is supra segmental phoneme.

Segmental phonemes can be identified either physically or auditory, in the stream of speech. Vowels and consonants are the segmental phonemes of a language. While, supra segmental phonemes cannot exist independently as they exist only with the segmental phonemes. Segmental phonemes constitute alphabets of a language.

III.   LITERATURE SURVEY:
One of the Major challenges before Information Technology is to overcome language barriers dividing humanity so that everyone can communicate with everyone else on the planet in real time. South Asia is one of those unique parts of the world where a single language is written

in different scripts. For Example, Consider the case of Punjabi language, spoken by tens of millions of people, but written in Indian East Punjab (28.2 million) 1 in Gurumukhi script (a Left to Right script based on Devanagari), in Pakistan West Punjab (62.6 million), in Shahmukhi script (a Right to Left script based on Arabic), and by growing number of Punjabis (2 million) in the EU and the US in the Roman script. Efficient IPA convertors translate the given text into its corresponding IPA text so that anyone can understand who know Punjabi with online IPA converter we can achieve this target. In future we will create online IPA convertor for Punjabi language for Gurumukhi script.

IPA convertors available for following regional languages are:

1. Assamese
2. Malayalam
3. Tamil
4. Marathi
5. Telugu
6. Bengali
7. Gujarati 8. Kannada

WHY NEED IPA?

Different letters have different sounds all over the world. In IPA, different sounds are put in symbols so that a word can be read in the same manner across the globe by the people of any region. It also provides different symbols for each sound of the alphabet used in the language. Example: cake (here the sound of c is a hard c) and in Cinema(here the sound of c is soft c) . Also, See and Sea have same pronunciation but in IPA these words will have different symbols so that these similar sound words can be differentiated accordingly.

IPA also provides different pronunciation of the same word in various dialects without possibility of confusion. It is standardized for everybody in every country.
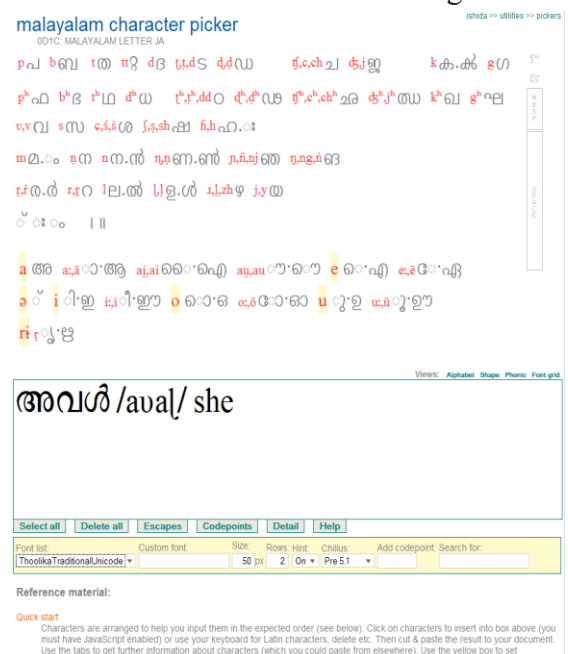
ONLINE IPA CONVERTORS IN VARIOUS LANGUAGES

1. ASSAMESE IPA CONVERTOR:

This language presently has a total of 11 vowel letters, that represent the 8 main vowel sounds of Assamese language, along with a number of vowel diphthongs. Vowel signs can be used in conjunction with consonants to modify the pronunciation of the consonant .

2. MALAYALAM IPA CONVERTOR :

In this language, an independent vowel letter is used as the first letter of a word that begins with

a

vowel. A consonant letter, despite its name, does not represent a pure consonant, but represents a consonant + a short vowel /a/ by default.

3. TAMIL IPA CONVERTOR :

Vowels can be either be short (1 mora) or long (2 morae). The long and short distinctions in Tamil include change in quality of the vowel as well the quantity (except «o»). For example, *«kaṇ<»* is /kəɳ/ but *«kāṇ»* is /kɑːɳ/ not /kəːɳ/. Similarly, *«cey»* is /sɛj/ but *«cēy»* is /seːj/ not /sɛːj/.

4. MARATHI IPA CONVERTOR:

Marathi language possesses some features of Sanskrit that have been lost in north-Indian Sanskrit-based languages such as Hindi and Bengali, especially when pronunciation of consonants and vowels is concerned. For instance, Marathi retains the original Sanskrit pronunciation of अं /əⁿ/, ऐ /əi/, and औ /əu/. However, as was done in Gujarati, Marathi speakers tend to pronounce ऋ ṛ somewhat similar to /ru/, unlike most other Indic languages which changed it to /ri/ (e.g. the original Sanskrit pronunciation of the language's name was *saṃskṛtam*, while in day-to-day Marathi and Gujarati it is *saṃskrut*. In other Indic languages it is closer to *sanskrit*). In speaking, Marathi allows for conservative stress patterns in words like राम (*rama*) with an emphasis on the ending vowel sound, a feature that has been lost in Hindi.
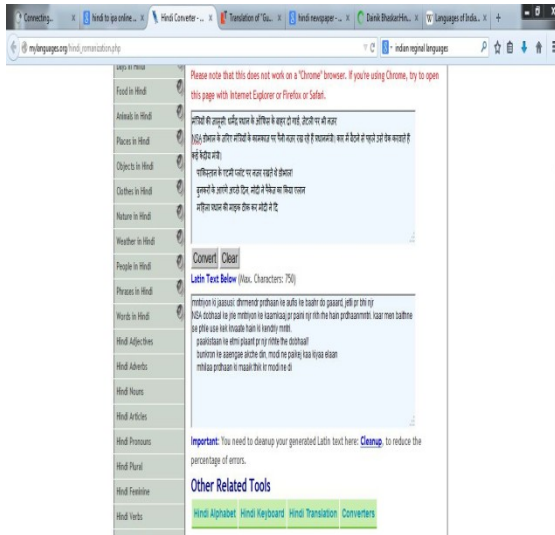
5. TELUGU IPA CONVERTOR:

Telugu words generally end in vowels. Telugu does not have contrastive stress, and speakers differ in the manner how they perceive stress. Most speakers judge it to be on the penultimate or final syllable, depending on word and vowel length.

6. BENGALI IPA CONVERTOR:

The phonetics of Bengali language consists of 29 consonants and 7 vowels, including 6 nasalized vowels. Bengali is widely known for its variety of diphthongs, combination of vowels occurring with the same syllable.

7. HINDI IPA CONVERTOR:

The convertor available for Hindi to IPA actually converts the Hindi text into Roman language. The IPA conversion is not available for this language.

PUNJABI PHONEMES:

Phonemes of Punjabi are of two types: Segmental and Super segmental categories. Stress, juncture, tone, intonation in Punjabi language are members of second type of Punjabi phonemes because of their dependence on segmental phonemes and presence over more than one segment.

Segmental phonemes are of the following types:

➢ Consonants: There are three types of conjunct consonants are written in which the altered form of the second consonant letter is sub joined to the first unaltered consonant letter.
➢ Diphthongs: The first member of a diphthong is always a short vowel and then it is followed by a long vowel.
➢ Geminates: Consonants are geminated by Addak .It is put on previous character and the next character is pronounced as full character and a half character.
➢ Prolative vowel: Addak is also used to elongate the long vowel when the vowel occurs at the end of a word. It makes the vowel one and half times the length of vowel.

SUPER-SEGMENTAL PHONEMES:

a) **Tone:** Punjabi is a tonal language. It uses five tonal characters and three types of tones which are classified as:

• **High tone:** It is a rising falling tone.
• **Low tone:** It is a rising falling tone. It is used in words using half only in the initial position.
• **Mid tone:** If a vowel does'nt bears any tone specification at the level of phonetic representation, it carries a mid tone.

b) **Stress:** Stress is not a prominent feature of Punjabi. It is utilized in disyllabic syllables to distinguish between the grammatical categories. In Punjabi language, accent is used on stressed syllables is a combination of length and pitch. Unstressed syllables lack length and a high pitch.

c) **Intonation:** A given sentence may be spoken in more than one way to present and express different altitudes, each of the different terms of the tonal system accordingly having more than one pitch exponent.

IV.   CHALLENGES:

TONAL MINIMAL PAIRS**:** Punjabi is a highly tonal language out of 22 languages that have been defined by constitution. Hence it poses extra challenges in building this language. The difference between mid tone, high tone, and low tone on similar minimal pairs requires sound phonetic knowledge for building the language.

STRESS**:** There are high gradations of small vowels according to their position of occurrence.

*HOMOPHONES***:** Words having same pronunciation but different grammatical category are called homophones.

WORD VARIENTS: Words having minor variation in spellings but having same meaning.

PUNJABI TO IPA CONVERTOR:

Available Punjabi to IPA convertor converts the given Punjabi text into its corresponding Roman text. There is one offline available convertor but it converts the written text into Roman text, rather than into IPA. Hence, in future I will work to convert the Punjabi text into its corresponding IPA, so that Punjabi becomes easy to speak by people all around the globe. In future we will work to create Punjabi to IPA convertor, so that people belonging to any region of the world can read Punjabi language written in Gurumukhi script without even knowing Gurumukhi.

REFERENCES:

[1]http://www.punjabiheritage.com/punjabi%20language.html

[2] http://www.gurmukhi.org

[3] http://www.sikhs.org/gurmukhi.

[4]http://mylanguages.org/hindi_romanization.php

[5]http://en.wikipedia.org/wiki/Help:IPA_for_Hindi_and_Urdu

[6] http://www.aksharangal.com/

[7]http://tdil-dc.in/index.php?option=com_up-download&task=view-download-tool&view=GET&toolid=216

[8] http://punjabi.changathi.com/

[9]http://en.wikipedia.org/wiki/Help:IPA_for_Tamil

[10] http://gujarati.changathi.com