ISSN 2319-2720
Volume 2, No.1, January – March 2013
International Journal of Computing, Communications and Networking
Available Online at http://warse.org/pdfs/2013/ijccn06212013.pdf

The Basic Grapheme to Phoneme (G2P) Rules for Bodo Language

Nabankur Pathak[1], Prof. P.H. Talukdar[2]
[1]Deptt of Instrumentation & USIC, Gauhati University, India, nabankur.pathak@gmail.com
[2]Deptt of Instrumentation & USIC, Gauhati University, India, phtassam@gmail.com

ABSTRACT

To develop a TTS system for a particular language, one must know the associated Grapheme to Phoneme rules. In a TTS system, the main task is the conversion of input text into linguistic representation, which is known as text-to-phonetic or grapheme-to-phoneme (G2P) conversion. The G2P conversion is highly depended on language. A very large set of different rules and their exceptions is needed to produce correct pronunciation and prosody for synthesized speech. This paper deals with the study and development of the Grapheme to Phoneme Rules for Bodo Language.

Key words : G2P, TTS, NLP, Schwa

1. INTRODUCTION

G2P refers to Grapheme to Phoneme conversion. This is the process of using rules to generate a pronunciation for a word. In another words the process of converting a sequence of letters into a sequence of phones is called grapheme-to-phoneme conversion, sometimes shortened G2P.

Formally, a Text to Speech (TTS) [1, 2] system converts a written text to speech or a sound file. The goal of a TTS system is to provide intelligible and natural speech. A TTS system consists of a Natural Language Processing (NLP) module and a Digital Signal Processing (DSP) module. The NLP module converts the text, that is the graphemes, to a string of phonemes [3,4]. It also encodes the intonation and prosodic information in the output string. In a Concatenative synthesis approach, the DSP module obtains the sound files from an acoustic inventory corresponding to the string of phonemes or diphones and concatenates them. Finally, it modulates the sound according to the intonation and prosodic information. Since in almost all languages, we hardly pronounce what we write, some linguistic analysis is necessary for the intelligibility of the speech.

In any TTS system, there are two basic stages. In the first stage it inputs text as input then converts it to phonetic string to be spoken. This stage is referred as Natural Language Processing (NLP). The second stage is the Speech Synthesis part in which speech is generated using a particular speech synthesis technique [5]. A major task in the Phonological Processing Component is to convert the input text into phonemic string using letter-to-sound rules [6]. For a natural language it is not possible to explicitly list all the words in that language [7]. So, when a new word comes up which is not explicitly listed, the lexical lookup fails in that case. TTS uses letter-to-sound rules to treat the new words which are out of the lexicon. So, letter-to-sound acts as a backup. Letter-to-sound rules system makes TTS light as it saves memory.

2. THE BODO LANGUAGE

The present Bodo language originated from the Sino-Tibetan family of language. The Sino-Tibetan family has many sub-branches, of which the Tibeto Burman had been living in the North-Eastern part of India. The Bodo sub group of Tibeto Burman family of language comprises of the following tribes: Kachari or Bodo, Tiwa, Dimasa, Garo, Rabha, Kakborok, Koch, Deuri and Moran. Bodo is the second largest major language of Assam. It is also the official language in the Bodo dominated area [8].

The Bodo language was introduced in the primary school levels of the Bodo dominated areas as medium of instruction. In the year 2005, the language got constitutional recognition as one of the 22 scheduled languages of India [9]. The Bodo language has got the status of one of the official languages of the state of Assam.

The Bodo phonemes consists of 6(six) vowels and 16(sixteen) consonants. Out of these 16 consonants 2(two) are semi vowel. They are as shown below-

a. Vowels : अ, आ, इ, उ, ए, औ
b. Consonants : ख, ग, ङ, ज, थ, द, न, फ, ब, म, र, ल, स, ह
c. Semi Vowels : य, व

3. GRAPHEME TO PHONEME CONVERSION TECHNIQUES

There are basically three broad categories [8] basically used for Grapheme to Phoneme (G2P) conversion. They are-

- Phonological Rule Based Approach
- Data Driven Approach
- Statistical Approach

3.1 Phonological rule based approach

The first technique is based on the assumption that the pronunciation of a letter or letter substring can be found if enough is known about its context. It uses handwritten phonological rules of the form

$$A[B]C \rightarrow d$$

This states that the letter substring B, in the context of A and C, rewrites as phoneme d. Since the mapping from orthography to sound is complex, especially for English, more than one rule is typically needed for the transformation. As described in [9], the conflicts that occur between rules as they are applied are resolved by keeping them in a set of sublists grouped by initial letter and ordered by the specificity of the rule. The most specific rule is placed at the top with more general rules towards the bottom. During transformation, the sublists are searched in a left to right process; for each letter, the appropriate sublist is searched from top to bottom until a match is found.

3.2 Data Driven approaches

Data Driven approach uses a model which maps graphemes to phonemes following automatic process by processing a training set. The operation of many data driven approaches is no different from the phonological rules; the questions stored in a decision tree could be reformulated as context-sensitive rules. In this approach, a letter to phoneme alignment is created automatically from the lexical database. The input word is then scanned for substrings that match the alignment database, and a pronunciation lattice is created. The decision function finds a possible pronunciation for the word by traversing the lattice and concatenating the phoneme labels on the arcs. The path chosen is shortest, and in the case of a tie, a heuristic is used.

3.3 Statistical Approach

It is also a kind of data driven technique, but it uses a Statistical methods to train a model from the data. The statistical methods mostly use Hidden Markov Model for G2P conversion.

4. METHODOLOGY FOR THE PRESENT STUDY

The basic main approach for the formulation of G2P (Grapheme to Phoneme) Rule for Bodo Language is the Hand Written Phonological Approach.

The formulation of the Grapheme rule the following steps have been followed-

i. Systematic analysis between the selected texts from the corpus and the recorded speech.
ii. Formulating the Hand written rule.
iii. Verification of the rule by a linguistic expert and testing of the rules in a HMM based speech synthesis system

5. GRAPHEME TO PHONEME RULES FOR BODO LANGUAGE

Like every other Indian languages, Bodo Language also shows some regular mapping from graphemes-to phonemes [7]. It has a systematic relationship between the written form of a word and its pronunciation. So, we preferred to write down letter-to-sound rules by hand.

Rules for Schwa Deletion or Retention:

The schwa is the vowel sound in many lightly pronounced unaccented syllables in words of more than one syllable. It is also known as neutral-vowel. Like every language Bodo has also schwa deletion or retention issue stated as follows-

i. The schwa of the first syllable is never deleted.
ii. If the word ends with the consonants like /v/(व) , then the associated schwa is retained.
iii. If /y/(य) comes after a syllable which consist a higher vowel like /i/(इ) or /u/(उ) than the associated schwa is retained.
iv. The schwa preceding a full vowel is retained to maintain lexical distinctions.
v. If the last syllable of the word contains a schwa and contexts 1 through 4 described above for the retention of the schwa do not occur, then the schwa is to be deleted.

General G2P Rule

These are the general grapheme to phoneme rules for Bodo language. These rule are followed while pronunciation of a Bodo word. They are-

i. If the Bodo vowel /w/(व) comes in between two consonants than it is pronounced as /ɯ/(ओ)
ii. If /n/(न) comes after any vowel at the end of a word than it will be pronounced as /ng/(न्ग)
iii. If /n/(न) comes after /o/(अ), then /o/(अ) will be pronounced as /u/(उ)
iv. If /j/ (ज) comes after /a/ (अ) or /e/ (ए), at the end of a word, then it will be pronounced as / ɯ / (ओ)
v. If /u/(उ) comes at the end of a word it is pronounced as /ɯ/(ओ)
vi. If a word consists the vowel /u/(उ) and /a/(अ) together than while pronouncing the word /w/(व) is pronounced in between them

It is also a kind of data driven technique, but it us

vii. If a word consists the vowel /i/(इ) and /a/(अ) together then while pronouncing the word /j/(ज) is pronounced in between them If a word ends with / ɯ/ (ओ) or /u/ (उ), then at the end /wa/(वा) is pronounced.

viii. If a word ends with /i/ (इ) or /e/(ए) then at the end of the word /ya/ (या) is pronounced

ix. If a word ends with /y/ (य), then at the end one more /ya/ (या) is pronounced.

x. If / ph/(फ) is associated with /w/(व); initially or finaly, /w/(व) is not pronounced

Nonsense Rule

A nonsense word is a syllable or group of syllables that can pronounced based on the phonetic rules of a language but which transmit no meaning to a reader or listener. Such words appear in a variety of literary contexts but generally only exist because of the sound of the nonsense word. In a speech synthesis process, detection of such cases is important for the naturalness of the synthesizer. The Basic nonsense rule for Bodo Language are-

i. No word in Bodo starts with / ŋ/(ङ), /v/ (व), /y/(य)

ii. No word in Bodo ends with /h/(ह)

iii. The diphthongs /ɔi/, /ui/, /ɯi/, /iu/ never occure at the starting of a word.

CONCLUSION

In this paper we have tried to discuss about the basic Grapheme to Phoneme rules for the Bodo Language. During the formulation of the rules Bodo corpora of about 10000 words has been used. The words are selected from continuous texts of Bodo newspaper, story-book, articles etc. All these rules are tested in a HTS (A HMM based Speech Synthesis Process). All these rules were able to produce 90% correct pronunciation of the Bodo words.

ACKNOWLEDGMENT

The authors Nabankur Pathak and Prof. P. H. Talukdar thanks Ministry of Communication and Information Technology (MIT), New Delhi, Govt. of India, for sponsoring the project on development of BODO TTS and giving financial support to the project.

REFERENCES

1. Dutoit T. **An Introduction to Text-To-Speech Synthesis**¸ *Kluwer Academic Publishers, 1996.*
2. Allen J. Hunnicut S. Klatt D. **From Text To Speech, The MITTALK System**. *Cambridge University Press, 1987.*
3. Hunnicut S. **Grapheme-to-Phoneme rules: a Review.** *Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden, QPSR 2-3, pp. 38-60.*
4. Belrhari R. Auberge V. Boe L.J. **From lexicon to rules: towards a descriptive method of French text-to-phonetics transcription,** *Proc. ICSLP 92, Alberta, pp. 1183-1186.*
5. T. Dutoit. **An Introduction to Text-to-Speech Synthesis**. *Kluwer Academic Publishers,Dordrecht, The Netherlands.1997*
6. Hussain, Sarmad. **Letter-to-Sound Conversion for Urdu Text-to- Speech System**. *In COLING 2004 Computational Approaches to Arabic Script-based Languages, Geneva, Switzerland. 2004*
7. Sarma Chandan. Talukdar P.H. **Dialect variation in Boro Language and Grapheme-to-Phoneme conversion rules to handle lexical lookup fails in Boro TTS System,** *International Journal of Scientific and Research Publications, Volume 2, Issue 9, September 2012*
8. Baro Madhu Ram. **Structure of Bodo Language**, *N. L. Publications.1990*
9. Basumatary Phukan. **An Introduction to the Bodo Language**, *Mittal Publication. 2005*