

Grapheme-To-Phoneme Tools for the Marathi Speech Synthesis

Sangramsing N. Kayte¹, Monica Mundada¹, Dr. Charansing N. Kayte²,
Dr. Bharti Gawali*

^{1,3}Department of Computer Science and Information Technology Dr. Babasaheb Ambedkar Marathwada University, Aurangabad

²Department of Digital and Cyber Forensic, Aurangabad, Maharashtra

ABSTRACT

We describe in detail a Grapheme-to-Phoneme (G2P) converter required for the development of a good quality Marathi Text-to-Speech (TTS) system. The Festival and Festvox framework is chosen for developing the Marathi TTS system. Since Festival does not provide complete language processing support specie to various languages, it needs to be augmented to facilitate the development of TTS systems in certain new languages. Because of this, a generic G2P converter has been developed. In the customized Marathi G2P converter, we have handled schwa deletion and compound word extraction. In the experiments carried out to test the Marathi G2P on a text segment of 2485 words, 91.47% word phonetisation accuracy is obtained. This Marathi G2P has been used for phonetising large text corpora which in turn is used in designing an inventory of phonetically rich sentences. The sentences ensured a good coverage of the phonetically valid di-phones using only 1.3% of the complete text corpora.

Keywords: Grapheme-to-Phoneme (G2P), TTS, Festival, festvox, di-phone, ICT.

I. INTRODUCTION

Marathi is an Indo-Aryan language spoken predominantly by Marathi people of Maharashtra. It is the official language and co-official language in Maharashtra and Goa states of Western India respectively, and is one of the 23 official languages of India. There were 73 million speakers in 2001; Marathi ranks 19th in the list of most spoken languages in the world. Marathi has the fourth largest number of native speakers in India. Marathi has some of the oldest literature of all modern Indo-Aryan languages, dating from about 900 AD. The major dialects of Marathi are Standard Marathi and the Varhadi dialect. There are other related languages such as Khandeshi, Dangi, Vadvali and Samavedi. Malvani Konkani has been heavily influenced by Marathi varieties, only a very small percentage of Indians use English as a means of communication. This fact, coupled with the prevalent low literacy rates make the use of conventional user interfaces difficult in India. Spoken language interfaces enabled with Text-to-Speech synthesis have the potential to make information and other ICT based services accessible to a large proportion of the population[1][2][3][18].

However, good quality Marathi TTS systems that can be used for real time deployment are not available. Though a number of research prototypes of Indian language TTS systems have been developed [2][3], none of these are of quality that can be compared to commercial grade TTS systems in languages like English, German and French. The

foremost reason for this is that developing a TTS system in a new language needs inputs for resolving language specific issues requiring close collaboration between linguists and technologists. Large amount of annotated data is required for developing language-processing modules like parts of speech (POS) taggers, syntactic parsers, and intonation and duration models. This is a resource intensive task, which is prohibitive to academic researchers in emerging economies like India [3][17][18][19].

In this paper, we describe the effort at HP Labs India to develop a Hindi TTS system based on the open source TTS framework, Festival [4][17][18][19]. This effort is a part of the Local Language Speech Technology Initiative (LLSTI) [2], which facilitates collaboration between motivated groups around the world, by enabling sharing of tools, expertise, and support and training for TTS development in local languages. LLSTI aims to develop a TTS framework around Festival that will allow for rapid development of TTS systems in any language.

The focus of this paper is on a Marathi G2P converter and on the design of phonetically balanced sentences. Certain linguistic characteristics of the Marathi language, like a large phone set and the schwa deletion issue, pose problems for developing unit selection inventories and Grapheme-to Phoneme converters for a Marathi TTS systems. Section 3 describes the modules and algorithms developed at HP Labs India to overcome these problems. In Section 4, the design of phonetically rich Marathi

sentences is presented. Though the effort is focused on Marathi, attention has been paid to make these modules as language independent as possible and scalable to other languages. The following section outlines the reason for choosing Festival as the base for developing such a system [2].

II. FESTIVAL FRAMEWORK

The Festival framework has been used extensively by the research community in speech synthesis. It has been chosen for implementing the Marathi TTS System because of its flexible & modular architecture, ease of configuration, and the ability to add new external modules. However, TTS system development for a new language requires substantial amount of work, especially in the text processing modules.

The language processing modules in Festival are not adequate for certain languages and the reliance on Scheme as a scripting language makes it difficult for linguists to incorporate the necessary language specific changes within Festival. Thus, we need new tools or modules to be plugged into Festival [4].

III. TEXT ANALYSIS MODULES

CREATED BY HP LABS INDIA

Grapheme-to-Phoneme (G2P) Conversion

In a TTS system, the G2P module converts the normalized orthographic text input into the underlying linguistic and phonetic representation. G2P conversion, therefore is the most basic step in a TTS system. It is preferable that the phoneme/phone sequence is also appropriately tagged with intonation markups.

Marathi, like most Maharashtra languages, uses a syllabic script that is largely phonetic in nature. Thus, in most cases, there is a one-to-one mapping between graphemes and the corresponding phones. Special ligatures are used to denote nasalization and homo-organic nasals. The Marathi phoneset consists of 86 phones due to the fact that it has a large set of distinct phonemes with a relatively small number of allophones. For example, the stop system in Marathi shows a four way distinction, viz., voiced vs. unvoiced, aspirated vs. unaspirated, for five places of articulation. There is also a phonological distinction between retroflex and alveolar, aspirated and non-aspirated taps. Also, each of the vowels has a phonologically distinct nasalized counterpart [17][18][19].

3.1.1. G2P Converter Architecture

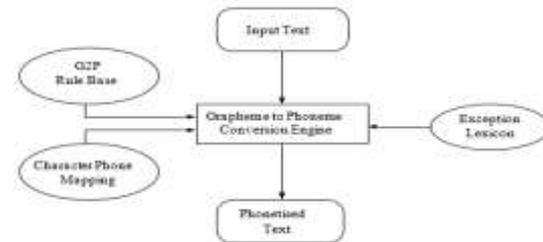


Fig. 1. G2P Converter Architecture

As part of the LLSTI initiative [4][16], a generic G2P conversion system has been developed at HP Labs India. The system consists of a rule processing engine which is language independent. Language specific information is fed into the system in the form of lexicon, rules and mapping. The architecture of the G2P is shown in Fig 1. This G2P framework, has then been customized for Marathi. The default character to phone mapping is defined in the mapping file. The format of the mapping is shown below and explained subsequently [4][5].

Character The orthographic representation of the character.

Type of the character, e.g. C (for Consonant), V (for Vowel) etc.

Class:-The class to which the character belongs. These class labels can be effectively used to write a rule representing a broad set of characters.

Phoneme The default phonetic representation of the Specific contexts are matched using rules. The system triggers the rule that best fits the current context. The rule format is shown below:

.....

Class label of the character as defined in the character phone mapping. Together these s represent the context that is being matched. Action specification node. Each such node has the form:

Action Type: Pos: Phoneme Str

Action Type This field specifies the type of this action node. Possible values are K (Keep), R (Replace), I (Insert) and A (Append).

Pos the index of the character which is covered by the context of the rule (1 Pos m). Phoneme Str Phoneme string used by this action node.

The G2P first looks, for each input word, in the exception lexicon. For the Marathi G2P, this lexicon is the phonetic compound word lexicon which is generated using the algorithm described. If the word is present in the lexicon, the phonetic transcription of the input word is taken from the lexicon itself. Otherwise, the G2P applies the rules on this word and produces the phonetic transcription.

Though Hindi phonology makes it easier to devise letter to-sound rules, there are certain exceptions to this. Some of these exceptions can be handled by straight-forward rules and some cannot. Schwa deletion is one such problem encountered in

Marathi G2P conversion. This problem arises with the phonetization of words containing the schwa vowel. Depending on certain constructs, the schwa vowel is deleted in some cases and retained in others. The orthography, however, provides minimal cues in recognizing these constructs. The following sub-section explains the issue of schwa deletion.

Schwa Deletion

The consonant graphemes in Hindi are associated with an inherent schwa vowel which is not represented in orthography. Let us consider the word kalam “pen”. The word is represented in orthography by the consonants viz. /k/, /l/ and /m/. The inherent schwa is associated with each of these consonants and the schwa is pronounced while pronouncing this word. As a rule of thumb in Hindi, a word ending schwa is always deleted. Vowels other than the schwa are explicitly represented in orthographic text with the help of specific ligatures or matras around the consonant.

Several phonological rules have been proposed in the literature [4] for schwa deletion. These rules take into account morpheme-internal as well as across morpheme-boundary information to explain this phenomenon. The Bell Labs Multilingual TTS attempted schwa deletion using morpheme boundary information [2]. The schwa is deleted in the context VC(C) __ CV where V stands for any vowel and C stands for any consonant. Presence of morpheme boundary on the left of the context for the schwa deletion block restricts application of these rules. The schwa is also not deleted if its deletion results in the formation of an invalid consonant cluster [17][18][19].

Encouraging results on schwa deletion using morpheme boundary information is reported in [6] [7]. Morpheme boundary in a word can be detected using a morphological analyzer. However, unavailability of high quality Hindi morphological analyzers prevents adopting this approach.

The following sub-sections describe how the schwa deletion problem has been handled in the Hindi text-to-speech system currently under development at HP Labs India.

Dhvani Schwa Deletion Rules

A set of schwa deletion rules have been incorporated in the Dhvani speech synthesis system [2] [8]. These rules are syllabic and can handle schwa deletion in independent words. Table 1 lists these rules. Here, W(i), C, V, H, Ch and ^e stand for the character of a word, consonant, vowel, health, consonant-health combination or a half consonant and schwa vowel, respectively. The rules are applied in a word from right-to-left.

Table 1. Dhvani Schwa Deletion Rules

W(n)	W(n-1)	Output
C	" "	C ^e
	V	CVC, if W(n-2) is a C; else VC
	C	C ^e C ^e C ^e
	Ch	C C ^e
Ch	" "	C
	V	CVC, if W(n-2) is a C; else VC
	C	C ^e C
	Ch	CHCV, if W(n) & W(n+1) are a CV pair & the corresponding 'half' sound is available; else C
V	" "	V
	V	V
	C	CV
	Ch	CV

A modified version of the schwa deletion rules are used in the Hindi G2P converter. These rules, however, fail in the case of compound words, which are highly prevalent in Hindi. Compound words are formed by two or more words concatenated to form a single word. To illustrate this point, let us consider the words lok “public” and sabhA “gathering”. These two are independent words. However, by combining these two words, the compound word loksabhA “lower house of the parliament” is formed. Application of the schwa deletion rules above on this word produces the output [l o k s bh A]^e which is incorrect.

This problem can be overcome if the constituents of a compound word are analysed separately, which requires a phonetic lexicon of compound words. Such a lexicon is not available for Hindi [2][8]. Hence, we have proposed an algorithm [9] to automatically extract compound words and identify its constituent parts from a large text corpus.

Compound Word Extraction

The proposed algorithm starts with the assumption that independently occurring words are valid atomic words. A trie-like [10] structure is used to store and efficiently match the words. A potential compound word is detected if the currently processed word is part of a word already present in the trie or a word in the trie is a sub-string of the current word. For each word (), there are several possibilities:

Case 1: () is neither in the trie nor a constituent of any of the words currently in the trie.

Case 2: () is an independent word in the trie.

Case 3: () is the initial part of a compound word currently present in the trie as an independent word.

Case 4: () is the second constituent of a compound word currently present in the trie as an independent word.

In Case 1, the algorithm will insert () into the trie. In Case 2, no change is needed. In Case 3, the algorithm will generate the second constituent () for each of the potential compound words where () is the

first constituent. After the generation of s, the end node corresponding to in the trie becomes a leaf node. The algorithm check for its presence in the trie for each of the generated s. If is present in the trie, the algorithm marks the combination () as a compound word. Otherwise, () is inserted into a suspicion list. The algorithm in its current form does not handle Case

4. Before processing each, the algorithm checks for its presence in the suspicion list. If is present in the suspicion list as the second constituent, the entry () is removed from the suspicion list and is marked as a compound word.

To illustrate the algorithm, let us consider a sample lexicon with words in the sequence lokgAthA, loksabhA, sabhA, lok, gAtha. If both constituents of a compound are present in the corpus, then the order of reading the compound and its constituent parts is irrelevant. After this the word lok is input. The algorithm first checks it in the suspicion list. If it is not found there, the word is matched against the contents of the trie generating the constituent forms gAthA and sabhA. sabhA is already present in the trie but gAthA is not yet read in. Hence, lok sabhA is marked as a compound word. But since gAthA is not present in the trie, lok gAthA is inserted into the suspicion list. After this, the algorithm processes the word gAthA. gAthA is present in the suspicion list as the second constituent. Hence, lok gAthA is removed from the suspicion list and is marked as a compound word. Then the algorithm tries to match gAthA in the trie and since it is not present, it is inserted into the trie.

Hence, at the end of the algorithm, the best possible atomic word forms are present in the trie with compounds decomposed into their constituent parts. An improvement of 1.6% in Hindi G2P conversion was observed as a result of incorporating the phonetic compound word lexicon into the Hindi G2P converter. A detailed description of the algorithm along with the issues involved are reported in [2][8].

G2P Results

Experiments were carried out to test the accuracy of the Hindi G2P. Details of the converter are presented in Table 2

Total Phones Used	86
Total Character-Phone Mappings Used	62
Total Number of Rules	84
Total Entries in Compound Word Lexicon	48428

Table 2. Customized Hindi G2P Details

To test the phonetisation accuracy of the Hindi G2P, a text file was randomly selected from the Emille corpus [2][4][8]. The text was phonetised using the G2P and the phonetic output was manually

verified. Phonetisation results are presented in Table 3.

Total Words Tested	3485
Wrong Schwa Deletion	81
Word Phonetisation Accuracy	97.67 %

Table 3. Result of Marathi G2P Conversion

IV. INVENTORY DESIGN

In concatenative speech synthesis approach, speech is generated by concatenating real or coded speech segments. A recent trend in concatenative synthesis approach is to use large databases of phonetically and prosodic ally varied speech, thereby minimizing the degradation caused by pitch and time scale modifications to match the target specification and to minimize the concatenation discontinuities at segment boundaries. The quality of output speech primarily depends on the quality of the speech corpus from which the speech segments are obtained. Consequently, the design of the speech corpus is a very important and a critical issue in the data-driven concatenative synthesis approaches. The corpus should contain as much variation as possible both phonetically and prosaically and at the same time the size of the corpus should be as small as possible both to minimize the overhead incurred in database preparation and to facilitate efficient unit selection at synthesis time. In our work, to generate optimal text to excise the speech corpus, a set of phonetically rich sentences are prepared.

The problem of selecting a set of phonetically rich sentences from a corpus can be mapped to the Set Covering

Problem (SCP). A survey of various algorithms for SCP is presented in [9]. The most widely used algorithms for SCP are greedy algorithms [2] [8]. The performances of greedy and spitting algorithms are comparable. We implemented greedy algorithm, since it is marginally better and less time consuming [10][11].

This algorithm builds the cover step-by-step. The algorithm recursively selects sentences. The first sentence is the sentence with the largest unit count. All units present in the sentence are then removed from the to-be-covered list of units. The sentence which covers maximum number of remaining units is selected next. This process continues until all the units are covered or the corpus is exhausted.

The units to be covered can be assigned weights based on their frequency in a corpus. If complete coverage is possible, then the weight of a unit can be made equal to the inverse of its frequency. This makes the algorithm focus more on less frequent units. The idea is that the high frequency units are anyway covered en-route. However, if it is known apriori that complete coverage is not possible, then the weight of the unit can be made equal to its frequency.

The Marathi G2P converter was used to phonetize the corpus. A di-phoneme frequency list was prepared by running CMU-Statistical Language Modeling Toolkit [12] [13] on the phonetised text corpus. An exhaustive list of 7392 diphones was generated using the Marathi phoneset. Phonotactic constraints of the language were not taken into consideration while generating the diphone list. Word beginning, word ending, sentence beginning & sentence ending contexts were considered while selecting the sentences. Silence-phone combinations were also taken into account. Text selection was carried out in three phases. Details of various corpora used in the text selection are presented in Table 4.

4.0.6. Phase 1 & 2

Sakale Newspaper text, which is part of the Emille corpus, was used in this step. From a total of 163,517 sentences, 665 sentences were selected resulting in the coverage of 2404 di-phones.

In the second phase, the idea was to cover the units not covered in the first phase. Samne newspaper text was used in this phase. From a total of 142122 sentences, 303 sentences were selected covering 379 di-phones. Hence, total di-phoneme coverage at the end of this step was 2783.

4.0.7. Phase 3

At the start of this phase, the three remaining corpora viz. Indian info, Miscellaneous (Misc.) and Spoken were analyzed to see their coverage of the di-phones which were not covered after the first two phases. The results are shown in Fig 2. It is clear from the graph that Marathi Misc corpus had the highest coverage. The Spoken corpus showed great promise

Corpus Name	Genre	Total Sentences	Total Words
Lokmat	Newspaper	163175	3393192
Sakale	Newspaper	142122	2910613
Samne	Newspaper	36057	607017
Marathi Misc.	Classified	128537	1728644
Spoken	Radio Text	17720	507952

Table 4. Marathi Corpora Statistics

in terms of coverage but the limited size of this corpus restricted its usage at this stage. Hence, the Misc corpus was used for text selection at this stage. Another interesting observation was the formation of the plateau with the addition of more newspaper text which is of the same genre as that of text used in Phases 1 & 2. This illustrates the fact that phonetic coverage attains saturation once a significant amount of text of one type is analyzed and additional text of the same type is unlikely to give significant increase in coverage.

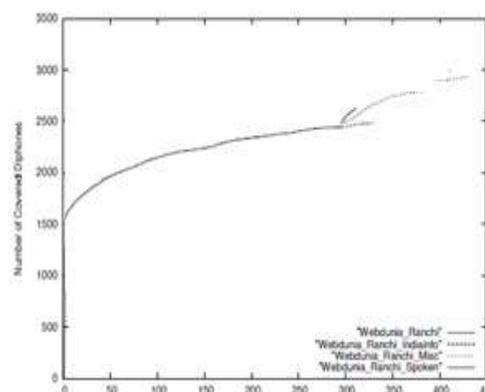


Fig. 2. Marathi Corpora Diphoneme Coverage Study

Another aspect observed during the analysis at this phase was that the covered-diphoneme to selected-sentence ratio was nearing one i.e. a complete sentence was being selected to cover just one unit. This is undesirable since the size of the unit inventory is directly proportional to the number of selected sentences and an unnecessarily big speech corpus results in additional overhead in speech-segmentation and increased search time in the unit selection stage. Hence, instead of the complete sentence, just the word containing the particular di-phoneme was included in the cover. This strategy selected 327 unique words resulting in a coverage of 373 di-phones [15]. Selected sentences were hand corrected to remove typographical errors and spelling mistakes. Final coverage results are presented in Table 5. The selected sentences and words are just 0.3% of the three text corpora (Webdunia, Ranchi Exp., Misc.) which were used in text selection.

Table 5. Results of Marathi Text Selection

Total Initial Di-phones	7392
Total Sentences Analyzed	433834
Count of Selected Sentences	968
Count of Selected Words	327
Total Di-phones Covered	3156
Total Uncovered Di-phones Count	4236

As a byproduct, the final uncovered di-phoneme list can be considered as phonotactically invalid Marathi di-phones since these di-phones remained uncovered after analyzing a large text segment of more than 400,000 lines. A separate list of words was prepared to cover the consonant clusters in Marathi [14][15][19].

V. CONCLUSION

An effective G2P conversion module has been developed for Marathi. This combines our new algorithm for automatically creating a compound word lexicon from a text corpus with the Dhvani

schwa deletion rule base. An inventory of phonetically rich Marathi sentences has also been created using a greedy algorithm on Marathi text corpora. By using a large text corpora from various sources, we are able to extract the phonotactic constraints in Marathi by identifying the forbidden di-phones.

REFERENCES

- [1] Sangramsing N.kayte "Marathi Isolated-Word Automatic Speech Recognition System based on Vector Quantization (VQ) approach" 101th Indian Science Congress Jammu University 03th Feb to 07 Feb 2014.
- [2] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis System for Hindi" International Journal of Advanced Research in Computer Science and Software Engineering -Volume 5, Issue 10, October-2015
- [3] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Di-phone-Based Concatenative Speech Synthesis Systems for Marathi Language" OSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 5, Ver. I (Sep -Oct. 2015), PP 76-81e-ISSN: 2319 -4200, p-ISSN No. : 2319 -4197
- [4] Sangramsing Kayte, Monica Mundada "Study of Marathi Phones for Synthesis of Marathi Speech from Text" International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-4, Issue-10) October 2015
- [5] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "A Review of Unit Selection Speech Synthesis International Journal of Advanced Research in Computer Science and Software Engineering -Volume 5, Issue 10, October-2015
- [6] "Dhvani - TTS System for Indian Languages," (<http://dhvani.sourceforge.net>), 2001.
- [7] Sangramsing Kayte, Monica Mundada, Santosh Gaikwad, Bharti Gawali "PERFORMANCE EVALUATION OF SPEECH SYNTHESIS TECHNIQUES FOR ENGLISH LANGUAGE " International Congress on Information and Communication Technology 9-10 October, 2015
- [8] Deepa S.R., A.G. Ramakrishnan, Kalika Bali, and Partha Pratim Talukdar, "Automatic Generation of Compound Words Lexicon for Hindi Speech Synthesis," Language Resources & Evaluation Conference (LREC), 2004.
- [9] Jeffrey D. Ullman Alfred V. Aho, John E. Hopcroft, Data Structure and Algorithms, Addison-Wesley Publishing Company, Reading, MA, 1983.
- [10] Emille corpus,"(<http://www.emille.lancs.ac.uk>), 2003.
- [11] Monica Mundada, Bharti Gawali, Sangramsing Kayte "Recognition and classification of speech and its related fluency disorders" International Journal of Computer Science and Information Technologies (IJCSIT)
- [12] A. Caprara, M. Fischetti, and P. Toth, "Algorithms for the Set Covering Problem," Tech. Rep. No. OR-98-3, DEIS-Operations Research Group, 1998.
- [13])Monica Mundada, Sangramsing Kayte, Dr. Bharti Gawali "Classification of Fluent and Dysfluent Speech Using KNN Classifier" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 9, September 2014
- [14] Sangramsing Kayte, Dr. Bharti Gawali "Marathi Speech Synthesis: A review" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 6 3708 - 3711
- [15] Monica Mundada, Sangramsing Kayte "Classification of speech and its related fluency disorders Using KNN" ISSN2231-0096 Volume-4 Number-3 Sept 2014
- [16] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte " Performance Calculation of Speech Synthesis Methods for Hindi language IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 13-19e-ISSN: 2319 -4200, p-ISSN No. : 2319 -4197
- [17] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "A Corpus-Based Concatenative Speech Synthesis System for Marathi" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 20-26e-ISSN: 2319 -4200, p-ISSN No. : 2319 -4197
- [18] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "A Marathi Hidden-Markov Model Based Speech Synthesis System" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 34-39e-ISSN: 2319 -4200, p-ISSN No. : 2319 -4197
- [19] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte "Implementation of Marathi Language Speech Databases for Large Dictionary" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 40-45e-ISSN: 2319 -4200, p-ISSN No. : 2319 -4197