Deriving Word Prosody from Orthography in Hindi

Somnath Roy
Centre for Linguistics
Jawaharlal Nehru University
New Delhi-110067
somnathroy86@gmail.com

Abstract

This study proposes a word prosody converter (WPC), which takes Hindi grapheme as input and yields output as a sequence of phonemes with syllable boundaries and stress mark. The WPC has two submodules connected in the linear fashion. The first submodule is a grapheme to phoneme (G2P) converter. The output of G2P converter is fed to the second submodule which is for prosody specific job. The second submodule consists of two finite state machines (FSMs). The first FSM does the syllabification and the second assigns prosodic labels to the syllabified strings. The prosodic labels are translated into the stressed and unstressed component using rules specific to the language. This study proposes a novel rule-based system which uses non-linear phonological rules with the provision of recursive foot structure for G2P conversion and prosodic labeling. The implementation[1] of the proposed rules outperforms the G2P models trained on the state of the art data-driven techniques such as joint sequence model (JSM) and LSTM.

1 Introduction

A dictionary is an essential component of a text-to-speech (TTS) and an automatic speech recognition (ASR) system. These systems are of open nature and can have an input word which is not present in the dictionary. Such input words are called out-of-vocabulary (OOV) words. Therefore, a G2P converter is required, which can generate the pronunciation of the OOV words. A G2P converter can be a rule-based or data driven system. A rule-based G2P converter relies on the expert knowledge (i.e., the rule-set designed by an expert). However, these rule-sets may not be exhaustive for capturing many language-specific properties such as word morphology and stress pattern (Pagel et al., 1998).Therefore, researchers nowadays rely on state-of-the art machine learning (data-driven) techniques for developing a G2P model. A data-driven system is trained using a manually annotated dataset. The manually annotated dataset contains words and its phonemic sequence. These datasets are language specific in nature. The machine learning algorithm learns the phonemic sequence for words based on the probabilistic or geometric calculation. These calculation varies across machine learning approaches. In data-driven approaches, one need not to worry about the language specific complexities such as word morphology and stress pattern. The algorithm automatically captures these patterns in the generated model. A data-driven G2P conversion process is broadly categorized into three subprocesses i) Sequence alignment ii) Model training and iii) Decoding (for details see (Novak et al., 2012)). Many data-driven techniques are available for G2P conversion. The important ones are decision tree (Black et al., 1998), Conditional Random Field (Wang and King, 2011), Hidden Markov Model (Taylor, 2005), Joint-Sequence techniques (Bisani and Ney, 2008) and Recurrent Neural Network (Rao et al., 2015).

The function of a word prosody model is similar to that of a grapheme to phoneme (G2P) converter. Moreover, it also describes syllable boundaries and predict stressed syllables in a word. The schematic diagram of word prosody model is shown in Fig 1. The accuracy of a word prosody module for Hindi language depends on an efficient solution of the two sub-problems well-known in Hindi phonology as schwa deletion and pronunciation of diacritic marks anusvara and anunasika (Ohala, 1983; Pandey, 1989; Pandey,

[1] https://github.com/somnat/Hindi-Word-Prosody-Hindi-G2P

S Bandyopadhyay, D S Sharma and R Sangal. Proc. of the 14th Intl. Conference on Natural Language Processing, pages 2–12,
Kolkata, India. December 2017. ©2016 NLP Association of India (NLPAI)

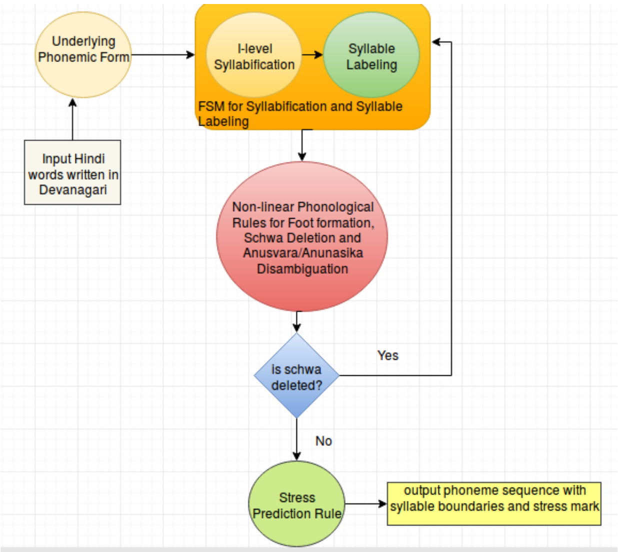
Underlying Phonemic Form
I-level Syllabification
Syllable Labeling
FSM for Syllabification and Syllable Labeling
Input Hindi words written in Devanagari
Non-linear Phonological Rules for Foot formation, Schwa Deletion and Anusvara/Anunasika Disambiguation
Is schwa deleted?
Yes
No
Stress Prediction Rule
output phoneme sequence with syllable boundaries and stress mark

Figure 1: Schematic Diagram of Word Prosody Model

1990; Narasimhan et al., 2004; Pandey, 2014). Ohala used linear phonological rules to derive surface phonemic form. Pandey showed the superiority of non-linear phonological rules over the linear one. The motivation for the current work is stated below.

a. In the past, Hindi G2P converters were implemented in the context of speech synthesis (Bali et al., 2004), (Narasimhan et al., 2004) and (Choudhury, 2003). However, these works have given partial attention to the anusvara/anunasika disambiguation. (Pandey, 2014) describes it as the problem of Hindi orthography.

b. These G2P converters are based on linear phonological rules proposed by (Ohala, 1983) with the exception of (Pandey, 2014). Non-linear phonological rules have advantages over linear one as explained below. (Bernhardt and Gilbert, 1992).

i. Non-linear rules capture both the prosodic and segmental information.

ii. The hierarchical representation used in non-linear framework captures more information; this results in a compact rule set.

c. Syllable is known to be a better unit for Hindi speech synthesis (Bellur et al., 2011; Kishore and Black, 2003). Therefore, a Hindi text-to-speech (TTS) system needs an automatic syllabification module. The automatic syllabification would be more useful if it could also predict the stressed syllables in words of natural speech as this would facilitate synthesis.

d. The usefulness of syllable as the basic linguistic unit in the context of speech recognition system has been explored in English (Ganapathiraju et al., 2001) and Tamil (Lakshmi and Murthy, 2006). Similar work for Hindi requires a software for syllabification. This work fulfills that need.

1.1 Main Contributions

- The WPC does not require the information of morphological boundaries. The proposed rules take into account the syllable patterns of compound, derived and inflected words.

- The syllabification and syllable labeling process follow finite state machine. The faultless syllabification and syllable labeling at underlying phonemic form yields better accuracy in schwa deletion and pronunciation of diacritic—anusvara and anunasika. The syllabification at underlying phonemic form is called as I-level syllabification in this work.

- The rules proposed in this study assume the extrametricality of foot unlike syllable as proposed in (Pandey, 2014). The contention is that the stress can be predicted elegantly using the notion of extrametrical foot (McCarthy and Prince, 1990; Crowhurst, 1994) . Also, the directionality is LR (left to right) unlike RL (right to left) used in (Pandey, 2014).

- Anusvara and anunasika are used interchangeably in Hindi. Therefore, both anusvara and anunasika is mapped to a hypothetical phoneme X at the underlying phonemic form. The decision for homo-organic nasal consonant or a nasalized vowel for phoneme X is based on the minimum moraic weight difference of the syllable having phoneme X and the next syllable. The moraic weight difference is calculated after schwa deletion and re-syllabification. The proposed mapping rule almost removes the pronunciation ambiguity related to anusvara and anunasika.

Rest of this paper is organized as follows. Section 2 describes the salient points of metrical phonology relevant to this work. Section 3 describes the process of syllabification and syllable labeling. Section 4 describes foot formation. Section 5 describes schwa deletion and re-syllabification. Section 6 describes the observations and rules for the anusvara and anunasika pronunciation. Section 7 describes the data-driven G2P systems implemented for Hindi. Section 8 compares the performance of current system to data-driven systems and previous rule-based implementations. Section 9 describes the rules for the prediction of the stressed syllables and reports the accuracy of current system for syllabification and stress prediction. The conclusion and limitations are written in Section 10.

2 Theoretical Background

Metrical phonology is based on nonlinear arrangement of the constituents of a phrase (Liberman and Prince, 1977; Selkirk, 1980; Hayes, 1980; Selkirk, 1986; Hayes, 1995; Apoussidou, 2006). The nonlinear arrangement is realized in the form of a tree with nodes as the constituents of a phrase. The constituents are syllable, foot, phonological word, phonological phrase and intonational phrase. Syllable is the lowest unit in the hierarchy dominated by foot, which in turn is dominated by a phonological word. The higher units such as phonological phrase and intonational phrase are not relevant in the current work (for clarity see fig 4 - 9). Syllable functions as a domain for segmental phonological rules. In non-linear phonology, the rules are written on the basis of interaction among syllables under the domain of higher constituents. A syllable has obligatory rhyme and optional coda. The syllables are also described by the moraic weight in quantity-sensitive languages such as Hindi (Pandey, 1989). Foot as a domain is used for describing stress and re-syllabification due to deletion of segment like schwa in languages such as French and Hindi. The foot is used as a musical meter and the concept is borrowed to non-linear phonology as a constituent (Selkirk, 1980; Hayes, 1995). Most of the quantity sensitive languages have binary foot, but some also allow degenerate foot. A binary foot is erected on either two syllables or on a single syllable having at least two moras. A single syllable having one mora, if projected as a foot, is called degenerate foot (Liberman and Prince, 1977; Hayes, 1995).Phonological word, also known as prosodic word, is a constituent unit of prosodic hierarchy above syllables or foot and below phonological phrases. Prosodic word is non-isomorphic to the grammatical word and the boundary of the former aligns with the morpho-syntactic boundary (Hall and Kleinhenz, 1999).

3 I-Level Syllabification

I-level syllabification is derived from the underlying phonemic form (UPF), which in turn is derived from orthography using the following mapping rules.

i. Each consonant in Devanagari script is inherently associated with the mid-central vowel called schwa or its lower counterpart ”a”[2].

ii. If a consonant is followed by a vowel diacritic mark, or a diacritic called halant, the inherent schwa is deleted.

iii. The inherent schwa is not realized in case of consonant at word final position.

iv. Two or three consonant together can form a ligature.

v. A short vowel at word final position is lengthened.

The following examples illustrate derivation of UPF from orthography:

/kml/ → kəməl (Lotus)
/kmAl/ → kəmaːl

The process of syllabification in Hindi was explored by (Ohala, 1983) and (Pandey, 1989; Pandey, 2014). Their analysis do not talk about the maximal onset principle for syllabification. The present analysis for syllabification follows maximum onset principle (Selkirk, 1984; Selkirk, 1981). The maximum onset principle is a sufficiency condition as demonstrated by the following examples.

i. मृत्युंजय (A name) → [mri] [tjun] [dʒəj]
→ *[mrit][jun][dʒəj]
→ *[mritj][un][dʒəj]
→ *[mrit][jundʒ][əj]

ii. कबूतर (Pigeon) → [kə] [bu] [tər]
→ *[kəb] [ut] [ər]
→ *[kə] [but] [ər]

In the above examples, the right hand side shows the potential syllable structures for a word. The square bracket denotes the syllable boundary. An asterisk before the syllable structure indicates that this potential syllable sequence is incorrect. The above examples show that either onsets are maximized or is equal to number of coda consonants for correct syllabification. It implies that the maximum onset principle hold for syllabification

[2] Hindi graphemes and its corresponding phoneme using international phonetic alphabet (IPA) and Roman symbols are described in Table 10.

in Hindi. Based on many such examples, following regular expressions are proposed for syllabification in Hindi.

i. v → [v]

ii. vv → [v][v]

iii. c*vcv → [c* v] [cv]

iv. c*vc1cv → [c*vc1] [cv]

v. c*vc1c2v → [c*v] [c1c2v]

vi. c*vc1c1v → [c*vc1] [c1v]

vii. c*vc2c2v → [c*vc2] [c2v]

viii. c*vcccv → [c*vc] [ccv]

In the above expressions, v denotes a vowel, ”c1” denotes a stop consonant, ”c2” represents a semivowel (r, l, v, j) and ”c” at intervocalic position denotes a consonant that is neither a stop nor a semivowel but can be any consonant at non-intervocalic position. An asterisk denotes the kleene star.

The finite state machine for the I-level syllabification is shown in Figure 2. It contains seventeen states with the start state as I and the final state as F. The orthography of an input word is transliterated into a sequence of consonants and vowels. The 8 syllabification rules are applied to this sequence to derive a symbol sequence in terms of c, c1, c2 and v. The symbol sequence is the input to the FSM in Figure 1. An arc between a pair of states in FSM is associated with a label (a pair of symbols seperated by ”/”). Suppose the symbol pair associated with an arc is ”x/y”. This indicates that whenever a symbol ”x” is fed to the state at the beginning of the arc, the system makes a transition along the arc and outputs the symbol ”y”. The label e/e symbolizes null input and null output for a transition. If part of a symbol string reaches to the final state F, then it is consumed, and a transition from F to I with arc label e/b takes place, where e is null and b denotes the syllable boundary of the consumed string. The remaining part of the string repeats the same process from initial state I until everything is consumed.

3.1 Syllable Labeling

Hindi is a quantity sensitive language. Therefore, syllables in Hindi are also described based on an attribute called syllable weight or moraic weight (Hayes, 1980; Pandey, 1989). A phonetic, phonological and typological description of syllable weight can be found in (Gordon, 2007). The following rules are used for label syllables based on syllable weights (for examples, see Table 1).

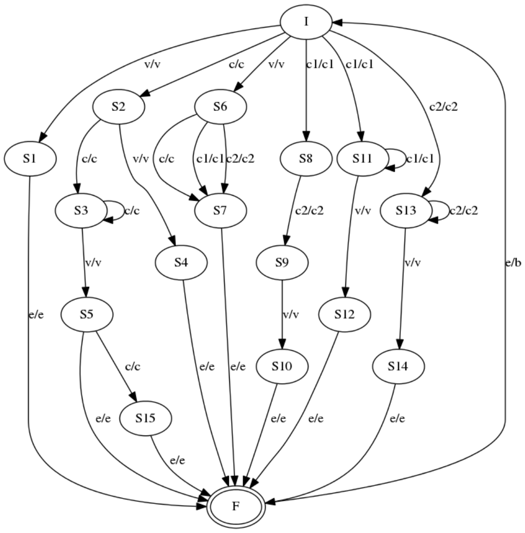

Figure 2: Finite State Machine for Syllabification

i. Each short vowel (like ə,u,i) and each coda consonant of a syllable are assigned a weight of one mora, while a long vowel (like aː,uː,iː) is assigned a weight of two moras.

ii. The syllables with one, two and three moras are called weak (w), heavy (h) and superheavy (sh) syllables respectively (Pandey, 1989; Hayes, 1989; Pandey, 1990).

Sylable Weight	Syllable Label	Gloss
1	[ki]^w	that
2	[mən]^h	soul
3	[gaːl]^{sh}	cheek

Table 1: Syllable labels according to syllable Weight

A finite state machine for syllable labeling is shown in Figure 3. The machine consists of one initial state (state I), seven non-final states and three final states (F1, F2 and F3). The syllabified string (i.e, the syllable boundary marked as b) is given as input to the initial state. Since, each short vowel gets one mora and long vowels get two moras, therefore, vowel type distinction is essential at syllable labeling stage. Each coda consonants get one mora and the onset consonants do not contribute to the moraic weight of syllables in Hindi. Therefore, consonant type distinction is not required at this stage. The symbol c, v_s and v_l stand for consonants, short vowels and long vowels respectively. The output symbol along an arc is either a syllable label or the reflection of the input itself. There is a null transition from each final state to the initial state so that the process can be repeated for the remaining part of the string. The FSM assigns the label ”w” to a syllable with phoneme sequences v_s, cv_s, ccv_s, ccc*v_s and ”h” to v_sc , c*v_sc, c*v_l, and ”sh” to c*v_scc, c*v_lc.

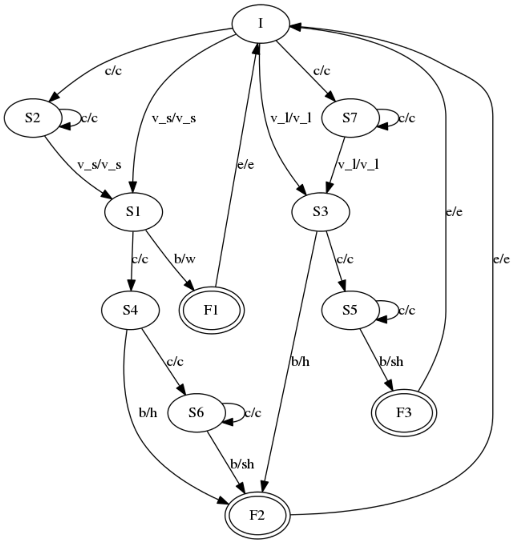

Figure 3: Finite State Machine for Syllable Labeling

4 Foot Formation

The concept of foot or feet is brought to linguistic from poetry. The notion of foot work as a metric to define stress pattern in a language (Jakobson, 1960). Moreover, foot also plays an important role in resyllabification due to deletion of a segment (Selkirk, 1996). In this section, a new approach of foot formation is described for Hindi. The approach is based on three assumptions and six rules. The rules apply in the direction from left to right.

4.1 Assumptions

i. Foot is formed using the labeled syllables of a word. The process of syllable labeling is described in the section 3.

ii. Foot is either binary branching or projected on at least a bimoraic syllable.

iii. A superfoot is formed either between a syllable and a foot, or between two foot.

4.2 Rules

The six rules for foot formation are listed below.

i. Weak to Weak Affinity Rule (WWAR): Two adjacent weak syllables form a binary foot and results into a bimoraic foot as shown in the Figure 4. The foot $<\sum_s>$is the extra metrical foot, which never bears any stress.

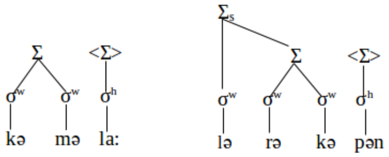

Figure 4: Bimoraic binary foot

ii. Heavy to Weak Affinity Rule (HWAR): A heavy and a weak adjacent syllables form trimoraic binary foot as shown in the Figure 5.

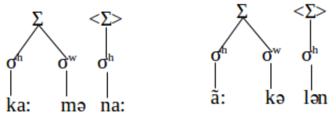

Figure 5: Trimoraic foot

iii. Weak to Heavy Affinity Rule (WHAR): This kind of foot is either formed in bisyllabic Hindi words or in loan words as shown in the Figure 6.

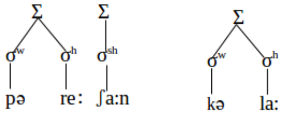

Figure 6: Trimoraic foot

iv. Heavy to Heavy Affinity Rule (HHAR): Two adjacent heavy syllables also form a binary foot as shown in the Figure 7.

v. Superheavy to Others Affinity Rule (SOAR): The superheavy syllables always projected as a foot as shown in Figure 7. The superfoot ($\sum_s$) is formed using one syllable and one foot. The projected foot constituent in $\sum_s$ could be either a new syllable after schwa deletion as shown in Figure 4 or a syllable which inherently bear stress i.e., the superheavy syllable as shown in Figure 8.

vi. List Affinity Rule (LAR): LAR is devised for handling words having same syllable structure at underlying phonemic form but realized differently at surface level. Such overlapping cases are stored in different list (usually different spreadsheets) and different foot formation rules are applied to these lists. The foot formation rules shown in Figure 9 describe four cases with overlapping syllable structure. These four cases represent four different list of words. These rules are called

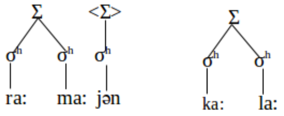

Figure 7: Heavy syllables forming a foot

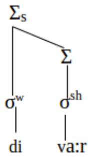

Figure 8: Superheavy syllables forming foot

as affinity hierarchy rules. The word affinity describes the interaction between different or similar type of syllables. The word 'hierarchy' is used because these rules apply in the order of their height. The top rule in the hierarchy applies first and so on. The decreasing order of height of these rules are LAR>WWAR >HWAR >WHAR >HHAR >SOAR.

5 Schwa Deletion and Resyllabification

Schwa deletion is an optional phenomena in Hindi. This means that schwa can be deleted or retained in the same environment, and the choice solely depends on the speaker and the context being used. Schwa deletion phenomena in Hindi helps speakers to utter a word quickly, i.e., the process of schwa deletion reduces the overall effort in terms of duration. It enables stress shift from one syllable to other. The following rules describe the contexts in which schwa gets deleted.

i. ə $\to \Phi/[\sigma^{\text{w}} - \sigma^{\text{w}}]_{\sum}$

ii. ə $\to \Phi/[\sigma^{\text{h}} - \sigma^{\text{w}}]_{\sum}$

iii. ə $\to \Phi/[\sigma^{\text{sh}} - \sigma^{\text{w}}]_{\sum}$

In other words, if the right most node of a foot is a weak syllable having schwa then schwa can be deleted.Application of the above schwa deletion (SD) rules and consequent re-syllabification of exemplar words are shown in Table 3. The process of schwa deletion and re-syllabification occur at foot level. In the process of re-syllabification, the bare consonant(s) after schwa deletion are assigned as coda consonant(s) to the preceding syllable. The foot structure for the examples in Table 2 can be found in the Section 4.

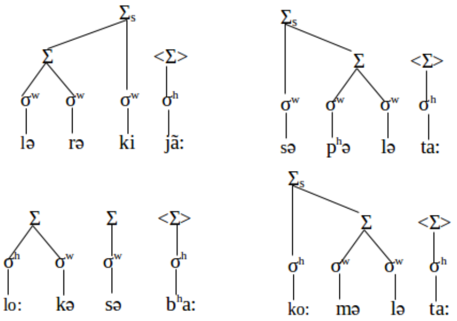

Figure 9: Examples of words with same syllable structure but different foot formation

UPF	SD	Resyllab	Gloss
kəməlaː	kəmlaː	[kəm][laː]	A Name
lərəkəpən	lərəkpən	[lə][rək][pən]	Childhood
kaːmənaː	kaːmnaː	[kaːm] [naː]	Wish
loːkəsəb$^{\text{h}}$aː	loːksəb$^{\text{h}}$aː	[loːk][sə][b$^{\text{h}}$aː]	Parliament
səp$^{\text{h}}$ələtaː	səp$^{\text{h}}$əltaː	[sə] [p$^{\text{h}}$əl] [taː]	Success

Table 2: Examples of word re-syllabification (Resyllab) after applying schwa deletion (SD) to underlying phonemic form (UPF)

6 Anusvara and Anunasika Pronunciation

Over the past, researchers have ignored the ambiguities in the pronunciation of anusvara/anunasika for G2P conversion. A simple finite state transducer for anusvara and anunasika is proposed by (Choudhury, 2003). (Pandey, 2014) says that the pronunciation ambiguities in anusvara/anunasika can be solved by preparing an exhaustive list of irregular cases. He further advocates the need for revision in the orthography to get rid of these irregular cases. However, not only in superscripted vowel diacritics as described in (Pandey, 2014), but in general, the present day Devanagari uses bindu and chandrabindu interchangeably. Some such examples are shown below in Table 3. The process of mapping anusvara and anunasika to appropriate phoneme is based on the observations and rules described below. The proposed approach maps both bindu and chandrabindu to the same phoneme X at the underlying phonological level. A single phoneme X for both anusvara and anunasika at underlying phonological level correctly captures the phonological structure of a word. The use of single phoneme X transforms the co-domain of the function G2P:A→ B and it becomes a one-to-one function from many-to-one function as shown in Fig. 10 and Fig.11 . The disambiguation rules apply on the phoneme X. Hence, these rules perform better for the words with identical use of bindu and chandrabindu. Moreover,The proposed approach uses both suprasegmental and segmental phonological constraints for anusvara/anunasika disambiguation.

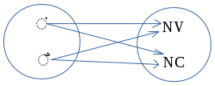

Figure 10: One-to-many mapping due to the identical use of anusvara and anunasika

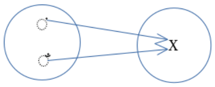

Figure 11: The use of phoneme X transform the codomain of G2P function. It becomes one-to-one function.

6.1 Observations

- Nasalization of vowel (NV) does not change the moraic weight of a syllable while the homo-organic nasal (HN) increases the moraic weight of a syllable by one unit.

- A syllable having phoneme for either anusvara or anunasika always tries to keep minimum moraic weigth difference with its succeeding syllable.

6.2 Rules

- Initially both anusvara and anunasika is mapped to phoneme, say, X. The moraic weight of X is assumed as one unit.

- The decision of using NV or HN is based on the comparison of moraic weight between the syllable to be mapped for NV/HN and the syllable succeeding to it.

  - If the moraic weight of the syllable containing phoneme X is greater than that of the next syllable, then replace the vowel and phoneme X by the corresponding NV.
  - If the moraic weight of the syllable containing phoneme X is less than or equal to that of the next syllable, then replace the phoneme X by the HN corresponding to the following phoneme.

- If the word final syllable contains the phoneme X at the last coda position then nasalize the vowel preceding X and delete X.

- If the word final syllable contains the phoneme X at non-final position and followed by a tʃ, tʃ^h, dʒ and dʒ^h then nasalize the vowel preceding X and delete X. Otherwise replace X by HN corresponding to the following phoneme.

Table 4 and 5 show examples of application of the above written rules for phonemic realization of anusvara and anunasika. In Table 4, the acronym Syllab denotes the syllable division and Mw denotes the moraic weight of syllables in the ordered pair.

7 Data Driven G2P Systems for Hindi

Two data-driven G2P models are trained on an expert annotated training lexicon of size 26454 words. These words are extracted from BBC Hindi. The first model is a joint sequence (JS) based G2P model trained using the sequitur (Bisani and Ney, 2008) toolkit. The second G2P model is a bidirectional deep LSTM model. The model configuration is same as reported in (Rao et al., 2015). Three forward and three backward hidden layers with 256 nodes at each layer is used. The output layer is a connectionist temporal classification (CTC) (Graves et al., 2006) layer and the error function is softmax.

8 Results

The publicly accessible Hindi wordnet (Bhattacharyya, 2010) is used for the testing purpose. The wordnet is first cleaned i.e., digits, hyphen and other special characters are removed. Long words especially compounds, derived and inflected words are picked up from different lexical categories like Noun, Verb, Adjective and Adverbs. The first list contains 3500 words for which schwa deletion rule applies at least once. A second list 700 words having diacritic for anusvara or anunasika. This implies that two test sets are used containing 3500 and 700 words. These sets are annotated by expert at three levels i) phonemic sequence, ii) syllable boundaries, and iii) stress mark. The G2P output for the

Graphemic Form-1	Graphemic Form-2	Correct Surface Form	Gloss
ऊँट	ऊंट	ũːʈ	Camel
कारवाँ	कारवां	kaːrvãː	coffle
कुँवारी	कुंवारी	kũvaːriː	Unmarried Girl
गाँधी	गांधी	gaːndhiː	A Name
गेहूँ	गेहूं	geːhũː	Wheat
घुँघरू	घुंघरू	g^{h}uŋg^{h} ruː	A Name
जाँघिया	जांघिया	dʒaːŋg^{h}ijaː	Underwear
जाएँ	जाएं	dʒaːẽː	Go(Honorific)
ढाँचे	ढांचे	ɖãːtʃeː	Shape
ताँबा	तांबा	taːmbaː	Copper
फँसा	फंसा	p^{h}ə̃saː	Trap (past)
भँवरी	भंवरी	b^{h}ə̃vriː	Loop
वर्षगाँठ	वर्षगांठ	vərʂgãː$ʈ^{h}$	Anniversary
शाहजहाँ	शाहजहां	ʃaːhdʒəhãː	A Name
हालाँकि	हालांकि	haːlãːkiː	However

Table 3: Examples of word in which diacritic mark for ansuvara and anunasika are used interchangeably at orthographic level but only one phoneme (i.e., either nasalized vowel or nasal consonant) emerges at surface phonemic form level

Grapheme	Syllab	Mw	Decision	Gloss
अंगूर	[aX] [guːr]	(2,3)	X=HN=ŋ	Grape
चींटी	[tʃiːX][ʈiː]	(3,2)	i:X=NV=ĩ:	Ant
अंबर	[əX][bər]	(2,2)	X=HN=m	Sky
अंधा	[əX][$ɖ^{h}$aː]	(2,2)	X=HN=n	Blind
आँचल	[aːX][tʃəl]	(3,2)	a:X=NV=ãː	A Name
सिंचाई	[siX][tʃaːiː]	(2,2)	X=HN=n	irrigation
जंजीर	[dʒəX][dʒiːr]	(2,3)	X=HN=n	chain
अंधेरे	[əX][d^{h}eː][reː]	(2,2)	X=HN=n	Dark
आंवले	[aːXv][leː]	(3,2)	a:X=NV=ãː	gooseberry

Table 4: Examples of applications of rules for phonemic realization of anusvara and anunasika

Grapheme	Phoneme	Gloss
गमलों	gəmlõː	Flowerpots
अनंत	ənənt	Infinity
धीरेंद्र	d^{h}iːreːndr	A name
पेंच	pẽːtʃ	Bolt
पाँच	pãːtʃ	Five

Table 5: Examples of applications of rule (Rule iii and iv) for phonemic realization of anusvara and anunasika

test sets by the proposed system, the data driven system and the previous systems are compared against the annotated test test. The rules of previous systems (Narasimhan et al., 2004),(Choudhury, 2003), (Bali et al., 2004) and (Pandey, 2014) are implemented in Python for comparison on the same test set. The example words where the others failed and current system succeded is shown in Table 9.The performance of these systems is reported below in Table 7. The present work has one limitation though.

> It cannot predict the stress pattern for the words having different part of speech categories as described in (Pandey, 2014) and (Dyrud, 2001). However, the number of such words in Hindi is small and can be listed. The future work will include a mechanism to predict stress for these words based on their part-of-speech category.

9 Prediction of Stressed Syllables

The process of resyllabification discussed in section 5 can upgrade syllables from weak to heavy or heavy to superheavy. Therefore, the after resyllabification the syllabified strings are fed to the FSM for syllable labeling. The labels assigned by FSM after resyllabification is called prosodic label. Table 6 shows examples of applications of syllable stress rule. Here a stressed syllable is preceded by a stress mark ('). The following rules translate the prosodic labels into stressed/unstressed component.

Type	**Stress Mark**	**Gloss**
w+h	ˈkəlaː	Art
h+h	ˈkaːlaː	Black
sh+h	aːˈraːm	Comfort
sh+sh	ˈraːmˈnaːt^{h}	A name
w+h+h	məˈhiːnaː	Month
sh+h+h	ˈaːlˈmaːriː	cupboard
h+h+sh	ˈhindusˈtaːn	Country Name

Table 6: Examples of applications of syllable stress rule

Systems	%Error1	% Error2
narasimhan et. al	9.57	12.08
choudhury	6.28	17.6
bali et. al.	5.2	8.27
pandey	7.5	9.4
JS Model	2.6	7.37
LSTM Model	2.16	3.56
Current System	0.45	1.5

Table 7: A summary of comparison of performance of the current system with previous rule-based systems and the state of the art data driven systems. The %Error1 and %Error2 denotes the word error rate due to schwa deletion and anusvara/anunasika pronunciation respectively.

i. Superheavy syllables are always stressed.

ii. Heavy syllables at the ultimate position are unstressed and stressed otherwise.

iii. Weak syllables at the penultimate position in bisyllabic words are stressed and unstressed otherwise.

The output of current system is evaluated for syllabification and stress prediction for the test set described in Section 8. The report is summarized in Table 8.

Testing Level	% Accuracy
Syllabification	100
Stressed Syllables	99.34

Table 8: A summary of testing of the current implementation for syllabification and prediction of stressed syllables

10 Conclusion

In this paper, a new approach for deriving Hindi word prosody is described. The WPC uses a novel rule-based G2P converter which outperforms the state of the art data-driven G2P systems and the previous rule-based system for Hindi. The proposed G2P system uses non-linear phonological rules with the provision of recursive foot. The proposed system has one limitation though. The system cannot predict the correct stress pattern of a word having two part-of-speech category as described in (Pandey, 2014). The proposed work can be further utilized in prosodic analysis by extracting stressed/unstressed syllables at textual level. The acoustic analysis can be performed by training the speech data and corresponding text using hidden markov model (HMM) or deep neural networks (DNN).

Grapheme	Other Systems	Current System	Gloss
लोकसभा	loːkəsb^{h}aː	loːksəb^{h}aː	Parliament
ताजमहल	taːdʒmhəl	taːdʒməhəl	A Name
कमलनयन	kəmlnəjən	kəməlnəjən	A Name
अनुसरण	anusrəɳ	anusərəɳ	To follow
अपवचन	əpəʋcən	əpʋəcən	Bad words
अपशकुन	əpəʃkun	əpəʃkun	Bad omen
बहुवचन	bəhuʋcən	bəhuʋəcən	Plural
उपग्रह	upəgrəh	upgrəh	satellite
हरभजन	hərəb^{h}dʒən	hərb^{h}ədʒən	A name
आंकने	aːŋkneː	ãːkneː	To Judge
क्योंकि	kjoːnkiː	kjõːkiː	Because
टाँग	ʈãːg	ʈaːŋg	Leg
पसलियां	pəsəlijãː	pəslijãː	Ribs

Table 9: Examples words where the others failed and current system succeeded

Acknowledgement

I would like to thank Pramod Pandey and Samudravijaya K for their comments and suggestions on the initial draft of this paper. I would also like to thank all the reviewers for their valuable comments.

References

Diana Apoussidou. 2006. *The learnability of metrical phonology*. Netherlands Graduate School of Linguistics.

Kalika Bali, Partha Pratim Talukdar, N Sridhar Krishna, and AG Ramakrishnan. 2004. Tools for the development of a hindi speech synthesis system. In *Fifth ISCA Workshop on Speech Synthesis*.

Devanagari	अ	आ	इ	ई	उ	ऊ	ए	ऐ	ओ	औ	क	ख	ग	घ
IPA	ə	aː	i	iː	u	uː	eː	æː	oː	əoː	k	k^h	g	g^h
Roman	a	aa	i	ii	u	oo	e	ei	o	au	k	kh	g	gh
Devanagari	च	छ	ज	झ	त	थ	द	ध	न	ट	ठ	ड	ढ	ण
IPA	tʃ	$tʃ^h$	dʒ	$dʒ^h$	t	t^h	d	d^h	n	ʈ	$ʈ^h$	ɖ	$ɽ^h$	ɳ
Roman	c	ch	j	jh	t	th	d	dh	n	tx	txh	dx	dxh	nx
Devanagari	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह	क्ष
IPA	p	p^h	b	b^h	m	j	r	l	ʋ	ʃ	ʂ	s	h	kʂ
Roman	p	ph	b	jh	m	y	r	l	v	sh	sx	s	h	ksh
Devanagari	त्र	ज़	क़	ड़	फ़	ङ	े	ँ	ं	ै	्र	ि	ी	matra उ ऊ
IPA	tr	z	q	ɽ	f	ŋ	eː	DMC	DMB	æː	RH	i	iː	u uː
Roman	tr	z	q	rx	f	ng	e	NV/NC	NV/NC	ei	r	i	ii	u oo

Table 10: Hindi grapheme and its corresponding phoneme in IPA and Roman. The DMC and DMB means Diacritic mark chandrabindu and bindu respectively. RH represents the Ra halant which is also a diacritic mark. u and oo in the last column represent matra (diacritic mark) for the vowel उ and ऊ respectively.

Ashwin Bellur, K Badri Narayan, K Raghava Krishnan, and Hema A Murthy. 2011. Prosody modeling for syllable-based concatenative speech synthesis of hindi and tamil. In *Communications (NCC), 2011 National Conference on*, pages 1–5. IEEE.

Barbara Bernhardt and John Gilbert. 1992. Applying linguistic theory to speech–language pathology: the case for nonlinear phonology. *Clinical Linguistics & Phonetics*, 6(1-2):123–145.

Pushpak Bhattacharyya. 2010. Indowordnet. In *In Proc. of LREC-10*. Citeseer.

Maximilian Bisani and Hermann Ney. 2008. Joint-sequence models for grapheme-to-phoneme conversion. *Speech communication*, 50(5):434–451.

Alan W Black, Kevin Lenzo, and Vincent Pagel. 1998. Issues in building general letter to sound rules.

Monojit Choudhury. 2003. Rule-based grapheme to phoneme mapping for hindi speech synthesis. In *90th Indian Science Congress of the International Speech Communication Association (ISCA), Bangalore, India*.

Megan J Crowhurst. 1994. Foot extrametricality and template mapping in cupeño. *Natural Language & Linguistic Theory*, 12(2):177–201.

Lars O Dyrud. 2001. *Hindi-Urdu: Stress accent or non-stress accent?* Ph.D. thesis, University of North Dakota.

Aravind Ganapathiraju, Jonathan Hamaker, Joseph Picone, Mark Ordowski, and George R Doddington. 2001. Syllable-based large vocabulary continuous speech recognition. *Speech and Audio Processing, IEEE Transactions on*, 9(4):358–366.

Matthew Gordon. 2007. *Syllable weight: phonetics, phonology, typology*. Routledge.

Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning*, pages 369–376. ACM.

T Alan Hall and Ursula Kleinhenz. 1999. *Studies on the phonological word*, volume 174. John Benjamins Publishing.

Bruce Philip Hayes. 1980. *A metrical theory of stress rules*. Ph.D. thesis, Massachusetts Institute of Technology.

Bruce Hayes. 1989. Compensatory lengthening in moraic phonology. *Linguistic inquiry*, 20(2):253–306.

Bruce Hayes. 1995. *Metrical stress theory: Principles and case studies*. University of Chicago Press.

Roman Jakobson. 1960. Linguistics and poetics. In *Style in language*, pages 350–377. MA: MIT Press.

S Prahallad Kishore and Alan W Black. 2003. Unit size in unit selection speech synthesis. In *INTERSPEECH*.

A Lakshmi and Hema A Murthy. 2006. A syllable based continuous speech recognizer for tamil. In *INTERSPEECH*.

Mark Liberman and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic inquiry*, 8(2):249–336.

John J McCarthy and Alan S Prince. 1990. Foot and word in prosodic morphology: The arabic broken plural. *Natural Language & Linguistic Theory*, 8(2):209–283.

Bhuvana Narasimhan, Richard Sproat, and George Kiraz. 2004. Schwa-deletion in hindi text-to-speech synthesis. *International Journal of Speech Technology*, 7(4):319–333.

Josef R Novak, Nobuaki Minematsu, and Keikichi Hirose. 2012. Wfst-based grapheme-to-phoneme conversion: Open source tools for alignment, model-building and decoding. In *FSMNLP*, pages 45–49.

Manjari Ohala. 1983. *Aspects of Hindi phonology*, volume 2. Motilal Banarsidass Publisher.

Vincent Pagel, Kevin Lenzo, and Alan Black. 1998. Letter to sound rules for accented lexicon compression. *arXiv preprint cmp-lg/9808010*.

Pramod Kumar Pandey. 1989. Word accentuation in hindi. *Lingua*, 77(1):37–73.

Pramod Kumar Pandey. 1990. Hindi schwa deletion. *Lingua*, 82(4):277–311.

Pramod Pandey. 2014. Akshara-to-sound rules for hindi. *Writing Systems Research*, 6(1):54–72.

Kanishka Rao, Fuchun Peng, Haşim Sak, and Françoise Beaufays. 2015. Grapheme-to-phoneme conversion using long short-term memory recurrent neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 4225–4229. IEEE.

Elisabeth O Selkirk. 1980. The role of prosodic categories in english word stress. *Linguistic inquiry*, 11(3):563–605.

Elisabeth O Selkirk. 1981. English compounding and the theory of word structure. *The scope of lexical rules*, pages 229–277.

Elisabeth O Selkirk. 1984. On the major class features and syllable theory.

Elisabeth O Selkirk. 1986. *Phonology and syntax: the relationship between sound and structure*. MIT press.

Elisabeth Selkirk. 1996. The prosodic structure of function words. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187:214.

Paul Taylor. 2005. Hidden markov models for grapheme to phoneme conversion. In *Interspeech*, pages 1973–1976.

Dong Wang and Simon King. 2011. Letter-to-sound pronunciation prediction using conditional random fields. *IEEE Signal Processing Letters*, 18(2):122–125.