Study of Indian English Pronunciation Variabilities Relative to Received Pronunciation

Priyanshi Pal[1(✉)], Shelly Jain[2], Chiranjeevi Yarra[2], Prasanta Kumar Ghosh[1], and Anil Kumar Vupalla[2]

[1] Electrical Department, Indian Institute of Science, Bengaluru 560012, India
priyanshipal53@gmail.com, prasantg@iisc.ac.in

[2] Language Technologies Research Center, IIIT Hyderabad, Hyderabad 500032, India
{chiranjeevi.yarra,anil.vuppala}@iiit.ac.in

Abstract. Analysis of Indian English (IE) pronunciation variabilities is useful in ASR and TTS modelling for the Indian context. Prior works characterised IE variabilities by reporting qualitative phonetic rules relative to Received Pronunciation (RP). However, such characterisations lack quantitative descriptors and data-driven analysis of diverse IE pronunciations, which could be due to the scarcity of phonetically labelled data. Furthermore, the versatility of IE stems from the influence of a large diversity of the speakers' mother tongues and demographic region differences. To address these issues, we consider the corpus Indic TIMIT and manually obtain 13,632 phonetic transcriptions in addition to those parts of the corpus. By performing a data-driven analysis on 15,974 phonetic transcriptions of 80 speakers from diverse regions of India, we present a new set of phonetic rules and validate them against the existing phonetic rules to identify their relevance. Finally, we test the efficacy of Grapheme-to-Phoneme (G2P) conversion developed based on the obtained rules considering Phoneme Error Rate (PER) as the metric for performance.

Keywords: Indian English · Pronunciation analysis · Received pronunciation · Phonetic rules · Phonetic rule validation

1 Introduction

India is a linguistically diverse country having more than 1,369 mother tongues [5]. The various languages spoken in India use a vast number of vowels and consonants [12]. Indian English (IE) pronunciation is affected by the varying influence of Indian native languages, which use many of these vowels and consonants. These variations pose a challenge in automatic speech recognition (ASR) and text-to-speech (TTS) synthesis systems in the Indian context. Consequently, these systems are rendered ineffective or yield performance degradation, which could be due to the inadequacy of labelled pronunciation data, which is lacking for Indian English speech [22].

A. Karpov et al. (Eds.): SPECOM 2023, LNAI 14338, pp. 339–349, 2023.
https://doi.org/10.1007/978-3-031-48309-7_28

[24] concluded that for better pronunciation modelling of a language that is non-native to the speaker, the characteristics of the speaker's native language must be considered in the modelling. Additionally, the differences in the phonemic inventory of various native Indian languages and English play a crucial role in a non-native Indian speaker's pronunciation of English phonemes. Typically, a non-native English speaker is inclined to map English phonemes to the closest phoneme in their native language [19]. As suggested in [13], a phoneme set developed to incorporate distinct characteristics of IE phonology can facilitate better pronunciation models for non-native speech. Approaches such as appropriate selection and optimisation of the phoneme set considered can increase the effectiveness of speech systems for non-native speech. [23] reported that speech recognition was more effective with phoneme set selection techniques for phoneme and word level speech recognition.

Considering these factors, there is a need to study IE pronunciations at the phonetic level to improve the speech systems for Indian speakers. Prior studies in the Indian context done to facilitate the adaptability of speech systems for non-native Indian speech are as follows. [1] reported phoneme selection rules for better naturalness and intelligibility in TTS for Marathi. [22] showed that certain IE accents are more recognisable than others, suggesting their suitability as canonical IE accents. [8] developed a linguistically-guided IE pronunciation dictionary for ASR by modifying the North American English (NAE) pronunciations in CMU (Carnegie Mellon University) Dictionary (often referred as CMUdict) [25] to IE using observed IE phonological features. For the few phonemes listed for comparison between NAE and IE pronunciations in IPA, the methodology to obtain phonological features of IE is unclear since ARPAbet is used in CMUdict. Hence, the peculiarities of IE obtained by comparing the canonical NAE pronunciations seem unsuitable. Other works in the Indian context have also studied phonetics and its influences, especially for particular Indian native languages. For instance, [16] examined Telugu speakers' L2 English phonetics.

Prior works lack approaches that focus on analysing sizeable datasets which are diverse in IE pronunciation variabilities using data-driven means. Typically, this results in capturing very few pronunciation variabilities in IE. Qualitative observations about various IE phonetic features can be informative; however, additional quantitative metrics can reveal the prevalence and significance of those observations. Furthermore, the data-driven rules are inherently dependent on the properties of the data used. In order to study the characteristics of IE using phonetic transcriptions, it must be ensured that the latter is reliable, consistent and representative of IE. This paper addresses these gaps by performing a data-driven analysis of phonetic transcriptions obtained by considering speech recordings in a linguistically diverse, Indic TIMIT corpus [27]. We gather existing qualitative phonetic rules relative to Received Pronunciation (RP) and report quantitative metrics to represent the prevalence of the phonetic features in IE and their probability of being representative of IE. We also present new rules found through our data analysis, which have not been discussed in the existing literature. Finally, we demonstrate the benefits of the obtained rules in building a Grapheme-to-Phoneme (G2P) conversion system for the automatic generation of IE pronunciations.

2 Data Annotation and Pre-processing

2.1 Indic TIMIT Corpus

We consider the speech data from Indic TIMIT [27] corpus for our work. In the corpus, 80 Indian English L2 speakers were considered from 6 regions of India, namely – North-East, East, North, Central, West, and South. From all these regions, speakers were recorded while speaking TIMIT stimuli [28], where each speaker was recorded for $2,342$ stimuli. The age of subjects ranged from 18-60 years. Cumulatively from all 80 speakers, a total of 240 h of speech data was obtained. From the considered 6 regions, a total of 5 groups were formed based on regions of the speakers' native language. The number of subjects in each group is 16 and they were gender balanced. The details of the groups are as described below:

Group 1 *(North East and East Regions)*: Maithili, Nepali, Oriya, Bengali, Assamese, Dimasa, Mog, and Manipuri.
Group 2 *(North and Central Regions)*: Malwi, Marwari, Punjabi and Hindi.
Group 3 *(West Region)*: Gujarati, Konkani, and Marathi.
Group 4 *(Upper South Region)*: Kannada and Telugu
Group 5 *(Lower South Region)*: Malayalam and Tamil.

The languages in these groups were identified based on their originating language families and also by considering how they are influenced by other language families. The languages in Groups 1, 2, and 3 originate from the Indo-Aryan language family, except for Dimasa, Mog and Manipuri, which are Tibeto-Burman languages. Assamese and Nepali are influenced by the Tibeto-Burman language family. Assamese and Bengali are also influenced by Austro-Asiatic language family. The languages in Groups 4 and 5 originate from Dravidian language family, wherein languages in the former group are also influenced by Indo-Aryan language family. The considered languages in these groups are spoken in proximate regions. Using information from these groups, further annotation is done. Since a large majority of the Indian population speak the languages considered in the corpus, subjects from these native languages were considered sufficient to cover the accent variabilities in IE.

2.2 Data Preparation

In Indic TIMIT, two linguists had transcribed a subset of the recordings of speakers that have native languages from all the 5 groups, totalling $2,342$. Apart from the pre-existing subset in the Indic TIMIT corpus, we collected annotations for $13,632$ recordings, totalling $15,974$ phonetic transcriptions for the analysis. They were phonetically transcribed sequentially into a total of 5 groups such that each group covered languages from all 5 region-based groups. This was done by considering one of the linguists who annotated a subset of transcriptions for Indic TIMIT Corpus. The linguist is affiliated with Spire Lab at the Electrical Engineering Department, Indian Institute of Science. We believe that the collected phonetic transcriptions could include the phonetic variations resulting from different native languages of the Indian population. A total of 108 IPA symbols were used for transcribing. The consistency of transcriptions was accessed by calculating Intra-Rater Agreement using Cohen's Kappa Score [6] for each group separately by repeating a sub-set of 200 files. The mean Cohen's Kappa Score was 0.827 across all groups, which indicates strong agreement. To perform an analysis of IE pronunciation, a pronunciation lexicon (containing words and respective phonetic pronunciations) was created considering the $15,974$ transcriptions from all 5 groups. The lexicon contains $16,664$ entries, each containing words and their corresponding pronunciation using IPA notation. Considering the existing literature in which IE pronunciation variations were described relative to RP, we also considered the RP canonical transcriptions obtained using BEEP pronunciation lexicon [17] to compare with IE for the analysis. The phone set of the BEEP lexicon is an extension of ARPAbet [18]. It was converted into IPA for comparison with phone-level IPA transcriptions of our speech data. The words (in the created lexicon from phonetic transcriptions) which contained "-" (ex: audio-visual) and were absent in the BEEP lexicon were added by considering the pronunciations of individual words already available in the lexicon. The phonetic transcriptions in our lexicon were mapped to that of the RP for the words common between our created and BEEP lexicons.

3 Indian English in Linguistic Literature

The influence of Indian native languages on the L2 English of Indian speakers attributes to the characteristic features of Indian English. Few linguistic works discussed these characteristics of IE relative to RP in the past, as mentioned in Table 1 and within this section. Considering these, we have assimilated the phonetic rules mentioned in the works. The phonetic rules based on English pronunciation, spoken by the Indian population regardless of their native language, are considered as *General IE Phonetic Rules* in Table 1. These features have been collectively described in the literature as characteristic identifiers of IE. The table also includes the phonetic rules which are specific to the native language of a speaker.

3.1 Context Dependent Phonetic Rules

Certain phonetic rules are based on the context, such as the position of vowels and consonants in a word. Found from literature, these are categorised into context dependent phonetic rules.

1. **Insertion or Omission of Phoneme:** In regions like Uttar Pradesh and Bihar, a short vowel /ɪ/ is prefixed at word-initial positions, as the following:

Table 1. Phonetic Rules mentioned in Literature.

General IE Phonetic Rules				Native Language Specific Phonetic Rules				
No.	RP	IE		No.	RP	IE	Native Language	References
1	/ɛ/	/e/ or /eː/	[3]	1	/ʃ/	/s/	Hindi, Telugu, Bengali, Bihari	[15,21]
2	/ʌ/	/ə/	[3]	2	/z/	/s/	Hindi, Telugu, Bengali, Bihari	[15]
3	/d/, /t/	/ɖ/, /ʈ/	[2,14,20,26]	3	/ɪ/	/i/	Assamese, Bengali, Bihari Hindi, Oriya	[3]
4	/θ/	/t̪ʰ/, /t̪/	[7,15]	4	/v/	/bh/	Bengali, Oriya, Assamese	[15]
5	/ð/	/d̪/	[7,15]	5	/ʒ/	/dʒ/	Kashmiri	[15]
6	/n/, /l/	/ə n/, /ə l/	[2]	6	/f/	/ph/	Gujarati, Marathi	[15]

"speech" becomes [ɪspiːtʃ] and "school" becomes [iskuːl] [11,14]. Few speakers add a semivowel before an initial vowel. Some examples would be, "every" ([jevri]), "about" ([jebauʈ]), and "old" ([woːlɖ]) [26]. Conversely, according to [14], sometimes people also tend to omit the semivowels /j/ and /w/. "Yet" is realized as [ɛʈ], "won't" as [oːnʈ].

2. **Rhoticity:** In words ending with the letter 'r', rhoticity is found in the IE pronunciation [26]. For example, as "letter" ends with /r/, it is realised as [ər]. However, whether IE is rhotic or non-rhotic is not unanimously concluded in the literature. [20] mentioned that although non-rhoticity is not governed by region, it is prevalent across regions.
3. **Monophthongisation of Diphthongs:** A majority of the Indian population uses monophthongs in their English, whereas diphthongs are used in RP [10]. For the diphthongs /eɪ/ and /əʊ/, the corresponding monophthongs /eː/ and /oː/ are used [2]. In certain contexts, such as word-final positions, these long vowels can be reduced to short vowels. For instance, in words like "today", these vowels are reduced to /e/ and /o/ [15]. In words similar to "near" and "square" where the vowel is succeeded by /r/ (i.e. /rV/), such as "period" and "area", IE generally uses /i/ and /e/ instead of /ɪə/ and /eə/ respectively [26].
4. **Word-specific Contexts:** 1) In the "-ed" inflexions which follow voiceless consonants, IE shows a greater use of /d/ over /t/. Some examples include words like "traced" as [treːsd] (IE) instead of [treːst] (RP), and "packed" as [pækd] (IE) in place of [pækt] (RP) [14].
   2) Double consonants in written English are often geminated. Few examples are: "matter" [mættər], "innate" [ɪnneːt], and "illegal" [ɪlliːgəl] [10].

4 Data Analysis

4.1 Procedure

In our data analysis, we aim to observe the variabilities of phonemes used in IE to those in RP. For this, we employ the many-to-many (m2m) aligner [9], which performs alignment followed by classification. Firstly, the phonemes in RP and IE pronunciation are aligned such that one or many phonemes of RP have the corresponding aligned IE phoneme(s) and vice versa. In addition to these alignments, we also obtain a confidence score from the m2m aligner indicating the likelihood between each set of aligned phonemes. We consider this confidence value (C.V.) for our analysis. Typically, m2m aligner is used for the prediction of phonemes, given graphemes. Therefore, the source is graphemes and the target is phonemes. In our analysis, we consider the source as RP phonemes and the target as IE phonemes for various words in the lexicon. We chose the maximum length specification in m2m aligner as 2 for obtaining alignments. The classification method provides C.V. for each aligned set of phonemes. Since C.V. indicates the likelihood of the IE phoneme(s) for a corresponding RP phoneme(s), we consider these values to validate the rules (phoneme mappings between IE and RP) obtained from the analysis based on aligned set of phonemes with the existing rules reported in the literature. We also consider the normalised frequency (N.F.) of occurrence corresponding to that rule, to indicate how recurrently it is observed. The frequency of occurrence of a rule is normalised by the total number of occurrences of RP phonemes of that rule.

The C.V. and N.F. both range from 0 to 1. The rules with C.V. of 0.10 and above are considered in this analysis. Furthermore, a minimum frequency of occurrence of 150 is also ensured for each rule to avoid C.V. and N.F. values derived from the low frequency of occurrence of the rule in the data. We grouped the phonetic rules into three categories in Table 2 based on their occurrence in literature, dataset, and as found using data-driven method:

- **Category 1** - *Phonetic rules mentioned in literature and observed in the dataset:* This contains IE phonetic rules, which were validated on the corpus by using data-driven methods.
- **Category 2** - *Phonetic rules observed in the dataset, but not discussed in literature:* This consists of phonetic rules which were observed with high C.V. and N.F.. However, discussion regarding them was not found in the linguistic works we studied.
- **Category 3** - *Phonetic rules mentioned in the literature but not observed in dataset:* The phonetic rules listed in this category have been discussed in the literature; however, they were either not present in our dataset or were not prominent enough to cross our thresholds for C.V. and frequency of occurrence. We also report the phoneme observed (Obs. IE) in our data in place of the expected phoneme (Exp. IE), which is mentioned in the literature and the C.V. and N.F. corresponding to Obs. IE.

4.2 Discussion

Category 1. The rules in rows 1–8 correspond to general IE features mentioned in the literature, were prominent in our dataset, indicated by high C.V. and N.F. values. For the rule in row 8, although for the phoneme /l/ in R.P, the most commonly observed corresponding phoneme in IE is /l/ in our data, there is a significant presence of usage of /ə l/ as well. Therefore, this phoneme

Table 2. IE Phonetic Rules relative to RP for all three categories. '*' Indicates Native language specific IE rules.

Category 1

No.	RP	IE	C.V.	N.F.
1	/ɛ/	/e/	0.917	0.912
2	/ʌ/	/ə/	0.94	0.932
3	/d/	/ɖ/	0.964	0.820
4	/t/	/ʈ/	0.964	0.851
5	/θ/	/t̪ʰ/	0.502	0.453
6	/θ/	/t̪/	0.45	0.381
7	/ð/	/d̪/	0.737	0.669
8	/l/	/ə l/	0.159	0.183
9(t)	*/z/	/s/	0.607	0.584
9(h)	*/z/	/s/	0.557	0.552
9(b)	*/z/	/s/	0.537	0.592
10	*/ɪ/	/i/	0.837	0.818

Category 2

No.	RP	IE	C.V.	N. F.
1	/ʊ/	/u/	0.980	0.747
2	/aʊ/	/au/	0.576	0.569
3	/j ʊ/	/u/	0.765	0.835
4	/ɝ/	/ə r/	0.866	0.525
5	/ɑ/	/a r/	0.624	0.237
6	/ɪ d/	/e ɖ/	0.912	0.373
7	/ʃ n/	/ə n/	0.843	0.893
8	/ə n/	/e n/	0.729	0.451

Category 3

No.	RP	Exp. IE	Obs. IE	C.V.	N.F.
1	/n/	/ə n/	/n/	0.873	0.902
2(t)	*/ʃ/	/s/	/ʃ/	0.402	0.375
2(h)	*/ʃ/	/s/	/ʃ/	0.336	0.334
2(b)	*/ʃ/	/s/	/ʃ/	0.47	0.420
3(t)	*/ʃ/	/s/	/ʃ ə/	0.508	0.416
3(h)	*/ʃ/	/s/	/ʃ ə/	0.481	0.399
3(b)	*/ʃ/	/s/	/ʃ ə/	0.461	0.389
4	*/v/	/b h/	/v/	0.964	0.942
5	*/f/	/p h/	/f/	0.984	0.984
6	/oʊ/	/oː/	/o/	0.925	0.731
7	/eɪ/	/eː/	/e/	0.953	0.727
8	/ɒ/	/ɔː/	/ɔ/	0.871	0.654

insertion happens sometimes, as mentioned in Table 1 under 3.1. For rules in row numbered 9, the native languages are Hindi and Bihari (h), Telugu (t) and Bengali (b). Row 10 is applicable for Hindi, Bengali, Assamese, and Oriya native languages are applicable. For these native language specific rules, only the transcriptions obtained from the native speakers of those languages are considered.

Category 2. The phonetic rule in row 1 has not been discussed in literature where comparisons between the RP /ʊ/ and IE /u/ have been made.

Consequently, it is possible that as a result of the phonetic rule in row 1, the rule in row 2 can be observed wherein, for the diphthong /aʊ/, /au/ is observed instead. There might also have been diphthongs where this replacement could be seen; however, such phonetic rules would not have met either the C.V. or minimum frequency of occurrence criteria in our dataset. Its influence can also be seen in the rule of row 3. However, the rule in row 3 also suggests the deletion of a phoneme. For example, as mentioned in Sect. 3 that in certain contexts, the semivowels /j/ and /w/ are omitted. Therefore, further investigations regarding the context of usage could be helpful in understanding the presence of this rule. The rules in rows 4 and 5 could suggest mild rhoticity in the speakers' accents, as mentioned in point 4 of Sect. 3. In the rule corresponding to row 6, the usage of retroflex /ɖ/ is clear from the validation of the rule in Category 1, row 3. However, there is little information regarding the presence of /e/ in /e ɖ/ or /ɪ/ in /ɪ d/.

Syllabification of /n/ and /l/ as /ə n/ and /ə l/ is discussed in [2]. The presence of rule in row 7 could indicate phoneme insertion, particularly in words ending with "-tion". For example, in the word "absorption", the RP pronunciation can be [əbsɔp**ʃn**], whereas [əbsɔp**ʃən**] can be the IE alternative. This may also be associated with the discussion in [15], where the insertion of /ə/ in a word-final cluster like "lm" in words such as "film" i.e./fɪləm/ is mentioned. In order to conclusively understand these phonetic rules, analysis of contexts along with native language is needed for the rules in rows 6, 7 and 8.

Category 3. Row 1 follows the description of syllabification [2]. However, unlike the schwa insertion in /ə l/ for /l, presence of /ə n/ wasn't observed for /n/. Instead, the prevalent usage was closer to RP phoneme /n/. The rules in rows numbered 2 and 3 share the same native languages as row numbered 9 in Category 1. Row 4 applies to Bengali, Oriya and Assamese speakers and row 5 for Gujarati or Marathi speakers. These are mentioned in rows 1, 4 and 6 in Table 1 under native language specific rules, as found in the literature. Following the rules in rows numbered 2,4 and 5, we observed the phonemes in IE (Obs. IE) of the corresponding native languages to be the same as RP. In row 2, apart from the occurrences where the IE phoneme /ʃ/ is the same as RP, we also observed /ʃ ə/ in IE, which is listed in row number 3. This could indicate the presence of /ə/ phoneme insertion.

In row 4, the prominent usage of the phoneme /v/ instead of /b h/ indicates a possibly vanishing /v/-/b h/ substitution. Similarly, row 5 indicates that /f/ was retained in its original form.

The rules in rows 6, 7 and 8 correspond to the diphthongs in RP, often substituted as monophthongs in IE. Contrary to the phonemes being substituted by a long vowel, we observed a wide usage of short vowels with high prominence. However, in certain contexts such as the ones which are mentioned in Sect. 3.1, IE often has short vowels substituting diphthongs.

Apart from the phonetic rules discussed above, the rules for the following phonemes /əʊ/, /ɛə/, /ɪe/ , /ɑː/, /ɔː/ couldn't be analysed as they are absent in the considered RP canonical transcriptions. Description for some of the rules related to them are as follows. In [3], the rule consists of RP phoneme /ɑː/ and its corresponding IE phoneme /aː/. Additionally, IE /ɒː/ is mentioned for RP /ɔː/ in another rule.

Finally, in addition to the rules mentioned in this category, with reference to point 5 under native-specific language features in Table 1, the phonetic rule specified for the Kashmiri native language is not analysed due to its absence in the languages considered in Indic TIMIT corpus.

Context Dependent Phonetic Rules. For the words ending with "-ed", the usage of IE /d/ instead of RP /t/ was barely observable. In our RP pronunciations for "-ed" ending words, instead of /t/, we observed /d/. This is contradictory to the description in 4. (Word-specific Contexts) under Sect. 3.1. Furthermore, the IE phoneme /ɖ/ for those words was observed in place of RP phoneme /d/, which is expected.

Many Indian languages have gemination in their verbal and orthographic forms, which explains the expectation for a native Indian language speaker to influence their L2 English similarly. However, a possible explanation for the absence of this behaviour in our data could be that speakers pronounced the correct phonetic sequence in the limited words where the context was applicable.

For gemination, we considered words with consonants such as "ll", "nn" and "tt". Very few instances of gemination by Indian speakers were observed. Lastly, we consider the insertion rule corresponding to /ɪ/ insertion as mentioned in point 1, particularly for words starting with "s". We consider Hindi speakers to validate /ɪ/ insertion. There were very rare instances where this was observed to happen. Apart from this, when word-initial positions were considered for semi-vowel insertion, the occurrences were very few.

4.3 Efficacy of G2P System Based on Phonetic Rules

We consider the Sequitur G2P conversion system [4] to show the effectiveness of the phonetic rules obtained from the proposed analysis. For the experimentation, we consider three pronunciation lexicons.

Table 3. Phoneme Error Rate (PER) for the lexicons.

Lexicon	IE	RP	IE_PRAG
PER	7%	47%	25%

The first one is referred to as IE lexicon, which is described in Sect. 2.2. It is constructed using unique pairs of words in the stimuli and their respective annotated phonetic transcriptions. Since the IE lexicon is obtained from phonetically annotated transcriptions for each word, the maximum performance can be achieved. Thus, IE lexicon can be considered as oracle lexicon. The second lexicon is referred to as RP lexicon, which is the BEEP pronunciation lexicon. Finally, the third one is referred to as IE_PRAG (**P**honetic **R**ule based **A**utomatically **G**enerated) lexicon, constructed from rules in Table 2 by substituting the phonemes of RP column in all the pronunciation sequences in the RP lexicon with the phonemes of IE column. Each substitution rule is applied to the fraction (equal to the N.F. in the table) of all possible candidates in the RP lexicon for the rule, chosen randomly. It is observed that the unique words vary in IE_PRAG, RP and IE lexicon. Thus, a similar approach mentioned in Sect. 2.2 is used to consider unique words common across all three. These are found to be a total of 6,720 out of which the pronunciation entries correspond to 5,376 (randomly chosen) words from all three lexicons for training the G2P system and the entries of the remaining words for testing. We consider Phoneme Error Rate (PER) as the metric for the evaluation on the test set. From the PER reported in Table 3 with all three lexicons, it is observed that the PER with IE_PRAG lexicon is lesser than that of the RP lexicon. This shows the benefit of the phonetic rules obtained from the proposed analysis for building a lexicon for IE automatically with G2P. Hence, IE_PRAG lexicon could be helpful in building better ASR and TTS in the Indian context.

5 Conclusion

Addressing the need to study and analyse IE pronunciation, we used a data-driven approach to explore the pronunciation variabilities of IE relative to RP.

For this, we phonetically transcribed 13, 632 utterances taken from the Indic TIMIT speech corpus. Considering a total of 15, 974 phonetic transcriptions, we presented a methodology to extract phonetic rules and validate them for their relevance and significance in the Indian context. We believed that the indicative rules helped determine relevant IE phonetic tendencies with higher confidence. Furthermore, we compared the performance of G2P conversion using lexicons constructed with and without the phonetic rules obtained in the proposed analysis. Further investigation is needed to analyse the quality of the new set of rules based on the influences from the native language-specific patterns. Additionally, inclusion of more annotators for phonetic transcriptions and their inter-annotator agreement can be presented with more data availability. Future directions include identifying the reasons for the absent rules reported in the analysis as well as further investigating the performance changes in ASR or TTS systems using the reported rules , along with incorporating other automatic G2P systems. Lastly, the examination of various phonetic combinations within speech signals remains a potential avenue for exploration.

References

1. Anil, M.C., Shirbahadurkar, S.D.: Phoneme selection rules for Marathi text to speech synthesis with Anuswar places. In: Thampi, S.M., Bandyopadhyay, S., Krishnan, S., Li, K.-C., Mosin, S., Ma, M. (eds.) Advances in Signal Processing and Intelligent Recognition Systems, pp. 501–509. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-28658-7_42
2. Bansal, R.K.: The pronunciation of English in India. Studies in the pronunciation of English: a commemorative volume in honour of AC Gimson, pp. 219–230 (1990)
3. Bansal, R.K., Harrison, J.B.: Spoken English: A Manual of Speech and Phonetics. Sangam (1994)
4. Bisani, M., Ney, H.: Joint-sequence models for grapheme-to-phoneme conversion. Speech Commun. **50**(5), 434–451 (2008)
5. Chandramouli, C., General, R.: Census of India 2011. Provisional Population Totals. New Delhi: Government of India, pp. 409–413 (2011)
6. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Measur. **20**(1), 37–46 (1960)
7. Gargesh, R.: Indian English: Phonology. In: A handbook of varieties of English, pp. 992–1002. De Gruyter Mouton (2008)
8. Huang, X., Jin, X., Li, Q., Zhang, K.: On construction of the ASR-oriented Indian English pronunciation dictionary. In: Proceedings of the 12th Language Resources and Evaluation Conference, pp. 6593–6598 (2020)
9. Jiampojamarn, S., Kondrak, G., Sherif, T.: Applying many-to-many alignments and hidden markov models to letter-to-phoneme conversion. In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, pp. 372–379. Association for Computational Linguistics, Rochester, New York (2007). https://www.aclweb.org/anthology/N/N07/N07-1047
10. Kachru, B.B.: English in South Asia, The Cambridge History of the English Language, vol. 5, pp. 497–553. Cambridge University Press (1994). https://doi.org/10.1017/CHOL9780521264785.011

11. Kachru, B.B.: Asian Englishes: beyond the canon, vol. 1. Hong Kong University Press (2005)
12. Kishore, S., Kumar, R., Sangal, R.: A data driven synthesis approach for Indian languages using syllable as basic unit. In: Proceedings of International Conference on NLP (ICON), pp. 311–316 (2002)
13. Kumar, R., Gangadharaiah, R., Rao, S., Prahallad, K., Rosé, C.P., Black, A.W.: Building a better indian english voice using" more data". In: SSW, pp. 90–94. Citeseer (2007)
14. Mesthrie, R.: Introduction: varieties of English in Africa and south and southeast Asia. In: 4 Africa, South and Southeast Asia, pp. 23–34. De Gruyter Mouton (2008)
15. Pingali, S.: Indian English. Edinburgh University Press (2009)
16. Pisegna, K., Volenec, V.: Phonology and phonetics of l2 Telugu English. Studies in Linguistics and Literature **5**, 46–69 (02 2021). https://doi.org/10.22158/sll.v5n1p46
17. Robinson, T.: Beep dictionary. BEEP dictionary (1996)
18. Robinson, T., Fransen, J., Pye, D., Foote, J., Renals, S.: Wsjcamo: a British English speech corpus for large vocabulary continuous speech recognition. In: 1995 International Conference on Acoustics, Speech, and Signal Processing. vol. 1, pp. 81–84. IEEE (1995)
19. Saikia, R., Singh, S.R.: Generating Manipuri English pronunciation dictionary using sequence labelling problem. In: 2016 International Conference on Asian Language Processing (IALP), pp. 67–70 (2016). https://doi.org/10.1109/IALP.2016.7875936
20. Sailaja, P.: Indian English: features and sociolinguistic aspects. Lang. Linguist. Compass **6**(6), 359–370 (2012)
21. Sirsa, H., Redford, M.A.: The effects of native language on Indian English sounds and timing patterns. J. Phon. **41**(6), 393–406 (2013)
22. Sitaram, S., Manjunatha, V., Bharadwaj, V., Choudhury, M., Bali, K., Tjalve, M.: Discovering canonical Indian English accents: a crowdsourcing-based approach. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (2018)
23. Vazhenina, D., Markov, K.: Phoneme set selection for Russian speech recognition. In: 2011 7th International Conference on Natural Language Processing and Knowledge Engineering, pp. 475–478. IEEE (2011)
24. Vignesh, S.R., Shanmugam, S.A., Murthy, H.A.: Significance of pseudo-syllables in building better acoustic models for Indian English TTS. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5620–5624 (2016). https://doi.org/10.1109/ICASSP.2016.7472753
25. Weide, R.: The carnegie mellon pronouncing dictionary [cmudict. 0.6]. Version 0.6. Available at [www. speech. cs. cmu. edu/cgi-bin/cmudict] (2005)
26. Wells, J.C.: Accents of English: Volume 3: Beyond the British Isles, vol. 3. Cambridge University Press (1982)
27. Yarra, C., Aggarwal, R., Rajpal, A., Ghosh, P.K.: Indic timit and indic English lexicon: A speech database of Indian speakers using timit stimuli and a lexicon from their mispronunciations. In: 2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA), pp. 1–6. IEEE (2019)
28. Zue, V., Seneff, S., Glass, J.: Speech database development at MIT: Timit and beyond. Speech Commun. **9**(4), 351–356 (1990)