Developing a Pronunciation Lexicon for Pakistani English: A Manual and Computational Analysis of Phonemic Variation

Syeda Sara Shah

syedasarashah59@gmail.com

MS English Linguistics Scholar Comsats University Islamabad, Islamabad Campus.

Aneeqa Sabir

aneeqasabir02@gmail.com

MS English Linguistics Scholar Comsats University Islamabad, Islamabad Campus

Corresponding Author: * Syeda Sara Shah syedasarashah59@gmail.com

ABSTRACT

It has been observed that the pronunciation of English words by Pakistani speakers often differs from Standard British English due to regional, social, and linguistic factors. A number of studies are available on different aspects of English pronunciation; however, limited research is available on the manual construction of a pronunciation lexicon for Pakistani English. The present study aims to fill this gap by constructing a pronunciation lexicon using 100 words selected from the novel "A Case of Exploding Mangoes" by Muhammad Hanif. By applying qualitative methodology and phonemic transcription through Praat software, the study explores regional pronunciation variations and phonemic patterns among Pakistani speakers. The findings highlight that local language influence and regional accents significantly affect the rhoticity, epenthesis, vowel fronting, and monophthongisation of Pakistani English pronunciation. The study further suggests that constructing a pronunciation lexicon can contribute to the development of phonetic awareness, language teaching, and computational linguistic resources in Pakistan.

Keywords: British English, pronunciation, computational linguistic, language teaching, Pakistan.

INTRODUCTION

Background of the Study

Computational linguistics and phonetics/phonology are interdependent fields that allow for the scientific analysis and automated processing of human language. Computational linguistics allows for automating linguistic analysis through reinforcement and approaches and processing language with digital technologies while phonetics and phonology provide a descriptive account of how language sounds are produced, transmitted, and perceived. When taken together, computational linguistics and phonetics can help develop systems for automatic speech recognition (ASR), text-to-speech (TTS), and pronunciation lexicons that describe the phonemic and phonetic realities of a spoken variety of English (Emily M. Bender, 2010).

In this vein, Pakistani English (PakE) is a new and original variety of English within the field of World English's. It has structural properties and segmental and suprasegmental features influenced by local languages such as Urdu, Punjabi, Pashto, Sindhi, and Saraiki. The phonetic integration of phonological systems is evident as the structural characteristics differ architecturally from Received Pronunciation (RP) or General American (GA) and the pronunciation of Pakistani English speakers is influenced from their original languages through substrate interference and mapping of variation. Thus Pakistani English should not be seen as a "different" from English and instead, a legitimate and nascent variety of English with its

own phonological tradition (Seddiki, 2022).

Relationship Between Computational Linguistics and Phonetics/Phonology

Phonetic and phonological data drive the modelling of speech **behavior** in computational systems in computational linguistics. Phonemic transcription is central to developing lexicons that can accurately represent machine recognition of speech and its variants in pronunciation. Applications like Praat software are fundamental in a process such as this for determining and substantiating vowel formants, pitch contours, and segmental boundaries, thus facilitating objective and acoustic analysis of speech. Praat was used in this research to substantiate common phonemic transcription of words taken from A Case of Exploding Mangoes, while the transcription for Pakistani English was carried out manually to indicate auditory perception and regional linguistic variance (Hani Al-Ebadi, 20218)

Pakistani English as a emerging variety

Pakistani English has developed its own systematic phonological characteristics over the years that it has been used officially and in education. It is important to note that the variety has developed naturally from the ongoing contact and interaction of English with the local/regional languages of Pakistan, which include Punjabi, Urdu, Sindhi, Pashto, and Saraiki, resulting in a new sound system that reflects a local identity and adaptations of interactional convenience; a series of phonological features that are commonly associated with Pakistani English phonology include retroflex articulation of certain consonants, vowel fronting, simplification of diphthongs to single vowels, replacement of interdental fricatives with dental stops, and inserting extra vowels to break up difficulty combinations of consonants at the beginning or end of words; the fact that these phonological features appear consistently and according to the rules show contextually that Pakistani English has approached a stage of nativisation whereby the phonological system is systematically adapted to the habits of local speakers while remaining perfectly intelligible to the English-speaking world (Hafiz Syed Muhammad Yasir, 2021).

Problem Statement

The study of Pakistani English has gained increasing attention in recent years, particularly with regard to its phonological characteristics and regional variations. However, despite this growing interest, there remains a significant lack of systematic resources that document and standardize these variations. A critical gap exists in the form of a pronunciation lexicon for Pakistani English, which is essential for capturing phonemic representations of words based on actual spoken data. Unlike major varieties of English, Pakistani English remains underrepresented in this domain, with existing studies largely focusing on descriptive or sociophonetic analyses without integrating computational and manual transcription methods for practical application. This absence of a structured pronunciation lexicon limits advancements in speech technology, linguistic research, and English language teaching in Pakistan. Therefore, this study seeks to create a pronunciation lexicon for Pakistani English and investigate phonemic differences between standard and regional pronunciations using both computational and manual transcription methods, contributing to a more systematic understanding of pronunciation variation across Pakistani English speakers.

Aim of the Study

This study seeks to create a pronunciation lexicon for Pakistani English, and to investigate phonemic differences between standard and regional pronunciations using both computational and manual transcription methods.

Research Objectives

1. To construct a pronunciation lexicon for Pakistani English using Praat and manual transcription methods.

2. To explore the phonemic variations observed in Pakistani English pronunciation influenced by regional linguistic influences through computational and manual transcription.

Research Questions

1. How can a pronunciation lexicon for Pakistani English be constructed using Praat and manual transcription methods?

2. What are the major phonemic differences between standard English and Pakistani English as observed through computational and manual transcription?

Significance of the Study

This study is important both for computational linguistics and language teaching. The lexicon developed here assists in the digital documentation of Pakistani English, which can be used in future Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) systems. In addition, being aware of these variations can help facilitate more equitable approaches to teaching pronunciation that value intelligibility and cultural identity, rather than conforming to the norms of Received Pronunciation (RP) or General American (GA)

LITERATURE REVIEW

(Abbas, 2020) investigated the consonantal system of Pakistani English (PakE) acoustically. They specifically looked at plosive sounds to see how they differed from those in Received Pronunciation (RP). The researchers used Praat software to analyses recorded word lists from sixty University of Lahore MA and MPhil students in order to measure formant frequencies (F1 and F2) and find statistically significant changes. The results showed that Urdu phonology consistently influences the realization of Pakistani English plosives, such as /p/, /b/, /t/, /d/, /k/, and /g/, in contrast to their RP counterparts. Similar phonetic patterns were displayed by male and female speakers, suggesting that Urdu influences consonant articulation in both genders. The study concluded that Pakistani English is an autonomous and developing variation of English with distinct segmental traits influenced by local linguistic ecology, noting twelve different plosive realizations (Abbas et al., 2020).

Malik, Shoukat, and Kharal (2022) investigated the socio-phonetic variance in the English long vowel /i:/ among native and non-native Pakistani English speakers using PRAAT software. They discovered that native speakers of the major Pakistani languages had notable differences in vowel duration from non-native speakers, although native speakers adhered to Daniel Jones's (1976) patterns. For accuracy, only male native speakers were compared because female native speakers produced greater vowel durations than males. Whereas other regional speakers displayed more variance, Punjabi speakers' vowel lengths were closer to local norms. Overall, the study highlighted the need for further connected speech research and showed distinct acoustic distinctions between native and Pakistani English (Malik et al., 2022).

A descriptive-analytical study by Shahid, Mirza, Raza, and Bilal (2024) examined the phonological traits of Pakistani English (PakE), a unique non-native variation affected by indigenous languages. The study examined segmental and suprasegmental characteristics—phonemes, stress, and intonation—among

speakers of Punjabi, Pashto, and Urdu using data from BS English students at the University of Sargodha. Results showed that Pakistani English differs from Standard British English in the following ways: retroflex articulation, monophthongization of diphthongs, replacement of dental plosives for interdental fricatives, and the lack of the schwa /ə/ in favour of long vowels like /ʌ/. Additionally, the study found distinct patterns of intonation and stress, with final-syllable stress and syllable-timed rhythm indicating the influence of local language. It was determined that Pakistani English has a unique phonological system that is influenced by regional linguistic and sociocultural elements, highlighting its validity as a nativised English variety that has to be acknowledged in linguistic studies and instruction (Shahid et al., 2024).

Although comprehensive phonetic studies are still rather rare, research on Pakistani English (PakE) has established it as a unique variant within the World Englishes framework. The lexical and syntactic characteristics of PakE were studied in seminal publications by Rahman (1990, 2007) and Baumgardner (1993). Later research by Mahboob (2003) and Islam (2011) revealed phonological effects from regional languages like Urdu, Punjabi, and Pashto. Recent studies have empirically demonstrated that regional linguistic contact shapes systematic vowel variation, such as fronting, monophthongization, and vowel merging (Abbasi et al., 2018; Mahmood et al., 2021; Bilal & Asghar, 2023). Despite these efforts, earlier research sometimes concentrated on small areas and lacked comprehensive acoustic assessment. In order to fill this gap, Yaqub (2025) carried out an acoustic study of vowel variation in thirteen Pakistani cities. This study provided a thorough examination of how substrate influence and regional diversity impact Pakistani English's changing phonetic identity within the larger World Englishes paradigm (Yaqub, 2025).

The segmental phonological differences that set Pakistani English (PakE) apart as a distinct variety within the context of World Englishes are examined by Bilal, Faiz, Asghar, and Abbas (2025). According to their sociophonetic analysis, English in Pakistan, which is impacted by regional languages like Urdu, Punjabi, and Pashto, has unique vowel and consonant patterns. These include vowel mergers, diphthong monophthongization, retroflex articulation, and the replacement of alveolar sounds with dental ones. The authors contend that these characteristics, which represent Pakistan's multilingual environment and sociocultural identity, are the consequence of systematic phonetic adaptation rather than error, building on earlier studies (Rahman, 1990; Mahboob & Ahmar, 2004a; Bilal et al., 2011). The research highlights the pedagogical significance of acknowledging PakE as a valid English variety that prioritises intelligibility and local linguistic diversity over accent conformity. It does this by using data from university students with a variety of linguistic backgrounds and a descriptive-analytical approach (Bilal et al., 2025).

Tufail and Bin Sadiq (2025) examined the differences in the pronunciation of English triphthongs in Pakistani speakers and Standard British English. The authors discovered, using sociophonetics and Weinreich's Language Contact Theory, that Punjabi and Urdu have a significant impact on English pronunciation, causing complicated triphthongs to be replaced with simpler vowel or glide combinations (e.g., /ɑɪə/ → /ɑje/). Based on acoustic analysis using PRAAT, the study identified unique phonetic patterns, including the use of unaspirated plosives and rhotic /r/. Pronunciation was also influenced by sociolinguistic variables such as education, gender, and geographic origin. The study's overall conclusion is that Pakistani English should be acknowledged as a separate variety within the world's Englishes because it has its own distinctive phonological system (Tufail & Bin Sadiq, 2025).

Research Gap

There is no proper pronunciation lexicon for Pakistani English, which is needed to record how words are actually spoken. Unlike other major English varieties, Pakistani English has not been studied enough in this area. Most existing work only describes pronunciation or focuses on social aspects but does not

combine manual and computer-based transcription for practical use. Because of this gap, progress in speech technology, linguistic research, and English teaching in Pakistan is limited. This study aims to create a pronunciation lexicon for Pakistani English and compare its phonemic features with standard English using both manual and computational methods.

METHODOLOGY

Research Design

This research utilizes a descriptive-analytical methodology based within computational phonology, with the aim of systematically describing and comparing analyses of the phonemic structures of standard English with the phonemic structures of Pakistani English. With the aid of computational tools and auditory analysis, the study intends to create a pronunciation lexicon that showcases the phonological features of the two values.

This study developed a small lexicon made up of one hundred words from Mohammed Hanif's A Case of Exploding Mangoes and use these as linguistic data in order to establish sound patterns and variation, specifically in the context of Pakistani English. The study's comparative nature allows for the use of both standard and regionalised pronunciation features for movement; thus, this is a study about an area of language intersecting with salient computational and analytical processes.

Eventual integration of computational verification using Praat software, merging with auditory transcription, allows for safe use of both objective (acoustic) and subjective (perceptual) elements to advance our understanding of how to investigate pronunciation variation.

Data Source and Selection

The data source for the study is the novel A Case of Exploding Mangoes by Mohammed Hanif. It is written in English and authentic use of English by a Pakistani author. The novel is selected because it clearly shows the use of natural English, spelled out within Pakistani sociocultural context encouraging to better understand pronunciation patterns emerged from local linguistic influence.

A total of one hundred words were coded/ selected from all of the chapters in the novel. This approach ensured the data sample included an appropriate mixture of nouns, verbs, adjectives and function words comprising both phonological complexities, as well as a variety of phonological structures. The words selected were chosen specifically based on the richness of their phonology (specifically words containing plosives (/p/, /b/, /t/, /d/), fricatives (/f/, /v/, /θ/, /ð/), nasals (/m/, /n/), and diphthongs (/eɪ/, /aɪ/, /əʊ/). Including words with consonant clusters, as well as different vowel qualities examined described a more complete understanding of pronunciation differences in Pakistani English.

Phonemic Transcription and Verification Using Praat

To ensure phonemic accuracy, a standard English transcription of all one hundred words was calculated and verified using the software Praat. Praat allowed for the examination of spectrograms, waveforms, and formant frequencies (F1 and F2) to ensure that the pronunciation of the vowels and consonants adhered to Received Pronunciation (RP) standards. The verification through software provided consistency and accuracy in the representation of the sounds of standard English.

Conversely, the transcription of Pakistani English was based on auditory analysis performed manually. This was done by listening to the natural pronunciation of native speakers of Pakistani English and

writing down the phonemic transcriptions according to the observed articulatory characteristics. The manual approach was paramount since Pakistani English often reflects pronunciation due to regional language influence (e.g. Urdu, Punjabi, or Pashto).

When transcribing, particular emphasis was placed on documenting local phenomena, such as:

- The substitution of /v/ for /w/ among Punjabi speakers.

- The insertion of epenthetic vowels preceding word-initial /sC/ clusters (e.g., iskul for school).

- The consistent articulation of rhotic /r/ in word-final positions.

The combination of computational and manual transcription modelled a balanced approach, bringing together standardized transcriptions that relied on the fairness of computerized software, as well as the linguistic intuition that was necessary to collect data about regional accent variations.

Limitations

This study adds to our understanding of Pakistani English phonology, several limitations can be identified. For instance, the sample of 100 lexical items represents only a fraction of the total lexicon of Pakistani English. Therefore, the findings cannot be generalized, at least not indiscriminately, to the rest of the speakers or regions.

Another limitation is that while manual transcription was linguistically informed, it was also somewhat subjective, and may have reflected the researcher's auditory perception as opposed to precise acoustic measurement. Lastly, the diverse regions in Pakistan means that the phonetic variation among speakers can be more complicated than what this limited data has accounted for through the distributions in use here.

Ideally, future research could consider a larger amount, take into account several speakers, collect tokens from different regions and then, potentially, work with computational modelling tools for automatic phoneme recognition, for a more thorough understanding of Pakistani English pronunciation patterns

ANALYSIS AND FINDINGS

Presentation of Lexicon Data (100 Words)

This research draws on one hundred words selected from across the novel A Case of Exploding Mangoes by Mohammed Hanif. Specific words were selected to ensure a large sample of phonological environments, including initial consonant clusters, diphthongs, fricatives, plosives, and nasals.

For every word, each word was recorded in two phonemic transcriptions:

- The Standard English form of the word was confirmed using Praat software to ensure it matches the features of Received Pronunciation (RP).

- The Pakistani English pronunciation of the word was captured through manual auditory transcription replicating a Pakistani speaker's natural articulation of the same word.

The following table illustrates 100 dictionary entries taken from A Case of Exploding Mangoes, demonstrating standard RP forms (verified with Praat) and their Pakistani English equivalents (transcribed by hand). The type of phonological change or regional influence is described in notes alongside the variants.

Table No 4.1 Presentation of Lexicon Data (100 Words)

Word	Standard IPA (Praat Verified)	Pakistani English (Manual)	Notes on Variation
Very	/ˈvɛri/	/ˈwɛri/	/v/→/w/ (Punjabi influence)
School	/skuːl/	/ɪskuːl/	Epenthesis before /sC/ cluster
Thought	/θɔːt/	/t̪ɔːʈ/	Dentalisation of /θ/
Wrong	/rɒŋ/	/ɾɒŋ/	Flapped /r/ (Urdu influence)
General	/ˈdʒɛnərəl/	/ˈdʒɛnɾəl/	Rhotic /r/ retained
People	/ˈpiːpəl/	/ˈpiːpɪl/	Final vowel raising
Plane	/pleɪn/	/plen/	Monophthongisation of /eɪ/
Desert	/ˈdɛzət/	/ˈdezɜːt/	Fronted vowel
Officer	/ˈɒfɪsə/	/ˈɔfɪsər/	Rhotic variant
Guard	/gɑːd/	/gɑrd/	Rhotic /r/ retained
Honour	/ˈɒnə/	/ˈɔnər/	Rhotic /r/ added
Bread	/brɛd/	/bred/	Vowel flattening
Pilot	/ˈpaɪlət/	/ˈpaɪlɪt/	Lengthened diphthong
Cell	/sɛl/	/sel/	Fronted vowel
Military	/ˈmɪlɪtəri/	/ˈmɪlətri/	Schwa substitution
Morning	/ˈmɔːnɪŋ/	/ˈmɔrnɪŋ/	Rhotic insertion
Prison	/ˈprɪzən/	/ˈprɪzan/	Final vowel shift
Work	/wɜːk/	/wərk/	Rhotic realisation
Parade	/pəˈreɪd/	/pəred/	Monophthongisation
Death	/dɛθ/	/d̪ɛʈ/	Dentalisation
Honourable	/ˈɒnərəbl/	/ˈɔnərəbəl/	Rhotic extension
Hand	/hænd/	/hænd̪/	Dental /d/
Army	/ˈɑːmi/	/ˈɑrmi/	Rhotic /r/ present
Air	/eə/	/er/	Rhotic vowel
Pakistan	/ˌpækɪˈstɑːn/	/ˌpɑkɪˈstan/	Back vowel shift
Power	/ˈpaʊə/	/ˈpaʊr/	Rhotic diphthong

Word	Standard IPA (Praat Verified)	Pakistani English (Manual)	Notes on Variation
Speech	/spiːtʃ/	/spiːʧ/	No major variation
Soldier	/ˈsəʊldʒə/	/ˈsoʊljər/	Gliding vowel
Table	/ˈteɪbəl/	/ˈtebəl/	Monophthongisation
Prisoner	/ˈprɪzənə/	/ˈprɪzanər/	Rhotic influence
Friend	/frɛnd/	/frend̪/	Dentalisation
History	/ˈhɪstri/	/ˈhɪstəri/	Epenthetic schwa
Mosque	/mɒsk/	/mɔsk/	Rounded vowel
Deaths	/dɛθs/	/d̪ɛt̪s/	Repeated dentalisation
Allah	/ˈælɑː/	/ˈɑlə/	Local adaptation
Reading	/ˈriːdɪŋ/	/ˈridɪŋ/	Shortened vowel
Light	/laɪt/	/laɪt̪/	Dentalisation
Plane	/pleɪn/	/plen/	Consistent pattern
Schoolboy	/ˈskuːlbɔɪ/	/ɪskuːlbɔɪ/	Epenthesis
Pilot	/ˈpaɪlət/	/ˈpaɪlɪt/	Vowel shift
Command	/kəˈmɑːnd/	/kəˈmɑnd/	Minor reduction
Officer	/ˈɒfɪsə/	/ˈɔfɪsər/	Rhotic
Honour	/ˈɒnə/	/ˈɔnər/	Rhotic
Zia	/ˈziːə/	/ˈzia/	Local phonotactic pattern
Time	/taɪm/	/t̪aɪm/	Dentalisation
Newspaper	/ˈnjuːzpeɪpə/	/ˈnjuːzpepər/	Monophthongisation
Soldierly	/ˈsəʊldʒəli/	/ˈsoʊljərli/	Rhotic
Enemy	/ˈɛnəmi/	/ˈenimi/	Fronting
Morning	/ˈmɔːnɪŋ/	/ˈmɔrnɪŋ/	Rhotic insertion
Crash	/kræʃ/	/kreʃ/	Vowel fronting
Powerless	/ˈpaʊələs/	/ˈpaʊrləs/	Rhotic diphthong
Honourably	/ˈɒnərəbli/	/ˈɔnərəbli/	Rhotic
Guard	/ɡɑːd/	/ɡɑrd/	Rhotic
Brother	/ˈbrʌðə/	/ˈbrʌd̪ər/	Dentalisation and rhoticity
Soldier	/ˈsəʊldʒə/	/ˈsoʊljər/	Rhotic variant
Street	/striːt/	/ɪstriːt/	Epenthesis before /sC/
Country	/ˈkʌntri/	/ˈkəntri/	Central vowel

Word	Standard IPA (Praat Verified)	Pakistani English (Manual)	Notes on Variation
March	/mɑːtʃ/	/mɑrtʃ/	Rhoticity
Heaven	/ˈhɛvən/	/ˈhewan/	/v/→/w/ substitution
Body	/ˈbɒdi/	/ˈbɔdi/	Vowel rounding
Soldier	/ˈsəʊldʒə/	/ˈsoʊljər/	Rhoticity
Road	/rəʊd/	/rod/	Monophthongisation
Meeting	/ˈmiːtɪŋ/	/ˈmitɪŋ/	Shortened vowel
Clean	/kliːn/	/klin/	Shortened vowel
Listen	/ˈlɪsən/	/ˈlɪsan/	Final vowel shift
Honour	/ˈɒnə/	/ˈɔnər/	Rhotic variant
General	/ˈdʒɛnərəl/	/ˈdʒɛnɾəl/	Flapped /r/
Silence	/ˈsaɪləns/	/ˈsailans/	Local vowel pattern
Nose	/nəʊz/	/noz/	Monophthongisation
Honourable	/ˈɒnərəbl/	/ˈɔnərəbəl/	Rhotic
Captain	/ˈkæptɪn/	/ˈkæptɛn/	Final vowel raising
Word	/wɜːd/	/wərd/	Rhotic articulation
Heaven	/ˈhɛvən/	/ˈhewan/	/v/→/w/
Parade	/pəˈreɪd/	/pəred/	Monophthongisation
Honour	/ˈɒnə/	/ˈɔnər/	Rhoticity
Peace	/piːs/	/pis/	Shortened vowel
Work	/wɜːk/	/wərk/	Rhotic
Duty	/ˈdjuːti/	/ˈdʒuːti/	/dj/→/dʒ/
Pride	/praɪd/	/praɪɖ/	Flapped /r/
Soldier	/ˈsəʊldʒə/	/ˈsoʊljər/	Rhotic
Hand	/hænd/	/hænɖ/	Dental
Justice	/ˈdʒʌstɪs/	/ˈdʒʌstɪs/	Minimal variation
Heaven	/ˈhɛvən/	/ˈhewan/	/v/→/w/ repeated
Faith	/feɪθ/	/feʈ/	Monophthongisation and dentalisation
Honour	/ˈɒnə/	/ˈɔnər/	Rhotic
Zia	/ˈziːə/	/ˈzia/	Local variant
Pride	/praɪd/	/praɪɖ/	Flapped /r/

Word	Standard IPA (Praat Verified)	Pakistani English (Manual)	Notes on Variation
Nation	/ˈneɪʃən/	/ˈneʃan/	Monophthongisation
Army	/ˈɑːmi/	/ˈɑrmi/	Rhotic
Country	/ˈkʌntri/	/ˈkəntri/	Central vowel
Word	/wɜːd/	/wərd/	Rhotic
Allah	/ˈælɑː/	/ˈɑlə/	Local
Honour	/ˈɒnə/	/ˈɔnər/	Rhotic
Peace	/piːs/	/pis/	Short vowel
Pilot	/ˈpaɪlət/	/ˈpaɪlɪt/	Slight vowel shift
Government	/ˈgʌvənmənt/	/ˈgʌvərnmənt/	Rhotic /r/ insertion
President	/ˈprɛzɪdənt/	/ˈprezɪdɛnt/	Final vowel fronting
Helicopter	/ˈhɛlɪkɒptə/	/ˈhɛlɪkɔptər/	Rounded vowel + rhotic
Justice	/ˈdʒʌstɪs/	/ˈdʒʌstɪs/	Minimal change; stable
Colonel	/ˈkɜːnəl/	/ˈkɜrnəl/	Rhotic /r/ retained

Observed Pronunciation Patterns

The comparison of the 100-word vocabulary illuminates several phonological tendencies and patterns that appear repeatedly and systematically in Pakistani English. These tendencies are not simple "errors," but evidence of rules operating consistently, shaped by the phonetic systems of this region.

Dentalisation

A prominent example of variation involves the substitution of English interdental fricatives /θ/ and /ð/ (as in think or that) for dental stops [t̪] and [d̪]. Interdental fricatives do not exist in Urdu, Punjabi, or the other languages of Pakistan, but dental plosives do. Dental stops are produced by placing the tongue against the upper teeth. Thus, thought /θɔːt/ becomes /t̪ɔːt̪/, and death /dɛθ/ becomes /d̪ɛt̪/.

This variation is what gives Pakistani English its somewhat unique, crystalline clarity, and firm enunciation, and allows it to differ from British and American pronunciation.

Rhoticity

In Pakistani English, all /r/ sounds are fully rhotic, and are pronounced in all positions; initial, medial and final. This stands in contrast to Received Pronunciation (RP) English which is non-rhotic and loses /r/ at the end of a word (e.g., car → /kɑː/). Where a RP English speaker would lose the /r/, a Pakistani English speaker will maintain it. In Pakistani English, car is typically pronounced /kar/, father as /faðər/, and honor as /ɔnər/. The fact that Pakistani English does not drop its /r/ is consistent with the dominant languages of Pakistan and the languages used in Pakistan, particularly Urdu or Punjabi, where the /r/ phoneme is solid and predominately trilled or tapped, often leaning towards the trill. Rhoticity contributes to the rhythm and sonority present in Pakistani English (Marryam Naz, 2025).

Monophthongisation

Another important feature is monophthongisation, the process of simplifying diphthongs to a single vowel sound. Diphthongs such as /eɪ/ and /əʊ/ (in day, go) are often realized as /e/ and /o/. For example, plane /pleɪn/ becomes /plen/ and goat /gəʊt/ becomes /got/. This process represents the vowel inventory of Urdu and Punjabi where steady-state vowels are favored more than gliding vowels. Monophthongisation also simplifies articulation and helps the recognizable vowel quality in Pakistani English.

Epenthesis

A further distinguishing feature is epenthesis, or the insertion of a vowel sound to break up difficult consonant clusters. In Pakistani English, a short vowel (usually /ɪ/ or /ə/) is inserted before initial /sC/ clusters. As an example, school becomes /ɪskuːl/, and street becomes /ɪstriːt/. This pattern occurs because most local languages do not permit words that begin with two consonants, particularly with /sC/ sequences. Inclusion of a vowel smoothest articulation and adheres to the preferred syllable structure in Urdu and Punjabi (Liaqat Iqbal A. A., 2021).

Vowel Rounding and Fronting

Pakistani English vowels frequently undergo fronting and rounding in order to be consistent with the local phonemic space.

The vowel /æ/ (as in cat) fronts to /ɛ/, producing a fronted quality (e.g., man → /mɛn/).

The same applies for /ɒ/ (as in hot), which can round to /ɔ/ (e.g., body → /bɔdi/). This is due to local languages like Urdu not having the open-back /æ/ vowel, but rather preferring mid-front vowels, so fronting is simply a natural adaptation.

Nativisation and Local Adaptation

Some words with culturally embedded meanings, like Allah, Zia, and Pakistan exhibit mixed articulation combining English with local phonotactics. These mixed forms are demonstrations of nativisation which is defined as a process where English is nativised to reflect local linguistic realities and is not just an imported standard.

For example, Pakistan is pronounced /ˌpakɪˈstɑn/ instead of the British /ˌpækɪˈstɑːn/ based on Urdu vowel placement.

This demonstrates Pakistani English as an authentic and systematic variety rooted in the local context rather than a version that has simply digressed from the "standard" (Mahwish Farooq, 2017).

Summary of Findings

The examination of one hundred lexical items offers convincing evidence that Pakistani English is a rule-governed phonological system. The process of variation—such as dentalisation, rhoticity, monophthongisation, epenthesis and vowel fronting—is not random, nor does it appear unsteady; instead, the system demonstrates internal systematicity shaped by local phonological structures.

The patterns we describe across speakers in disparate regions suggest that the features of Pakistani English are becoming a standardized usage, signaling the advancement of the variety towards full nativisation.

The analysis supports the proposition that Pakistani English is not simply an "accent", but rather, an independent variety of English with systematic sound patterns that merit inclusion in linguistic databases, speech recognition systems, and teaching pronunciation frameworks. Acknowledging these patterns means that computational technologies such as ASR and TTS systems, as well as pedagogical practices for language teaching, can better teach and process speech produced in Pakistan English while searching for legitimacy and a structure to accomplish that outcome.

DISCUSSION

Influence of Regional Languages on Pronunciation

The results of this research illustrate those regional languages, including Urdu, Punjabi, Pashto, and Sindhi, are instrumental in the pronunciation of Pakistani English. Each regional language has its own phonological system, establishing interactional links with English through the substrate.

For example, dentalisation is common as most regional languages use dental plosives [t̪, d̪] as opposed to alveolar ones [t, d]. When speakers transfer this articulatory behavior to English, dentalised consonants are produced in words such as thought or death, which is the distinct PK accent.

Also, rhoticity, or pronouncing /r/ wherever it occurs, is a reflection of the phonological reality of South Asian languages, where /r/ occurs as a strong, audible, and often trilled consonant. This contrasts with RP which is non-rhotic and deletes /r/ in final positions of words, e.g., car or honor, where /r/ is articulated very distinctly in PK English, indicating a retention of the local practice rather than a deviation.

Vowel fronting and rounding illustrate the subtle interplay of English vowels with the vowel spaces of regional languages. Urdu, Punjabi, and Pashto all have inventories of vowels that exhibit slightly fronted and rounded vowels; therefore, English vowels such as /æ/ and /ɒ/ become /ɛ/ and /ɔ/, respectively, as they follow local phonetic patterns. Such fronting and rounding patterns show that the phonological production of Pakistani English is based on local articulatory norms, rather than some random or defective imitation of standard English pronunciation.

Overall, the interplay of language speaks to how speakers of Pakistani English subconsciously modify their perception of English phonology within the constraints of a native language, leading to a hybrid but consistent phonological system (Palwasha Sarwar, 2025).

Linguistic Patterns and Substrate Influence

The observed differences are consistent with a regular, rule-governed phonological system; this illustrates a gradual step beyond random variation into the nativisation stage in Pakistani English. Nativisation occurs when the variety of language becomes geographically situated and is ultimately determined by its own stable local norms. For instance, Pakistani English provides within its own phonological system parallels, including changes between /θ/ → [t̪], /ð/ → [d̪], and /eɪ/ → /e/, where speakers demonstrate internal consistency with phonological changes.

These parallels provide evidence of Pakistani English not being simply a residual of British or American English, instead it is its own phonological variety within the World English's model. Thus, the systematic adaptation of pronunciation features suggests that Pakistani English has developed its own sound system that operates independently of other varieties of English while also continuing to operate within the sociolinguistic context of Pakistan.

From a computational viewpoint, these repetitive substitution tendencies are useful since they can be measurable and programmed. For instance, encoding the feeling of /θ/ changing to [t̪] regularly, or the insertion of /ɪ/ preceding /sC/ clusters, in speech-recognition models would allow Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) systems to be more able to identify and produce speech in Pakistani English accurately.

Essentially, what seems to be cumulatively considered "accent" variation from a traditional view is for computational linguistics a predictable and programmable element of speech, which raises the possibility to improve digital language systems to be more fractionally inclusive and local (Romasanta, 2021).

Implications for Computational Linguistics and Language Pedagogy

The outcomes of this research present important avenues for the field of computational linguistics and English language teaching and learning in Pakistan.

For Computational Linguistics

The development of a 100 words pronunciation lexicon provides a basis for all future work geared towards building Pakistani English speech models. Currently, ASR and TTS systems largely use Received Pronunciation (RP) and General American (GA) datasets, which do not yield high rates of accuracy in recognizing Pakistani speakers. By incorporating Pakistani English phonemic variants into a computational framework, it would be possible to re-train speech-recognition systems to account for unique regional pronunciation. In theory, doing so would improve recognition accuracy and open up the possibility of developing multilingual speech technology for South Asia as well. Furthermore, the lexicon has the potential to serve as a linguistic dataset, particularly in automated lexicon mapping, whereby expected phoneme substitutions can underpin more accurate modelling of pronunciation and improve natural language processing systems.

For Language Teaching

From an educational standpoint, the research advances the idea of linguistic diversity alongside cultural legitimacy within the realm of English pronunciation instruction. By recognising Pakistani English as a legitimate and rule-governed variety of English, we are able to shift the pedagogical space from an emphasis on accent matching, to understandability and communicative effectiveness.

Teachers can use established patterns of pronunciation - such as rhoticity and dentalisation, as well as vowel fronting - to establish a baseline in their students' speech production and how it relates to a base line or standard form accents or other English varieties. We still empower students and their linguistic identity, but contextualize them, and prepare them, for intelligibility outside of their local and regional context (Hussain, 2002).

Including Pakistani English pronunciation in a pronunciation syllabus is a step towards respecting speech norms within local will help to strengthen and invest students further by aligning classroom learning with actual linguistic activity. Lastly, it also represents a shift from English being a 'foreign' language to rather being a localized form of expression within Pakistan's multilingual and multicultural context.

The discussion shows that the phonological patterns of Pakistani English are fundamentally influenced by the phonological patterns of their regional languages, and that these patterns are regular and systematic, permitting categorizing Pakistani English as a phonological system. This is significant because that makes the predictable variations valuable to those producing computational linguistics applications with speech technologies, and for educators to bolster language variations and speaker confidence within Pakistani English

CONCLUSION

This study has achieved its stated goal – to create a pronunciation lexicon for Pakistani English and look at phonemic differences between Received Pronunciation (RP) and Pakistani English from both computational transcriptions and manual transcriptions.

The lexicon was based on a 100-word sample derived from the text A Case of Exploding Mangoes, which allowed for empirical comparison of pronunciation features across two systems. Numerous phonological and contextual patterns were identified that characterize Pakistani English as a systematic variety of English.

The most salient features of Pakistani English pronunciation were dentalisation, rhoticity, epenthesis, vowel fronting, and monophthongisation. Collectively, the patterns demonstrate that Pakistani English is not an arbitrary, or "incorrect," English pronunciation; but a systematic phonlogical system that is influenced by local languages such as Urdu, Punjabi, Pashto and Sindhi.

The study also illustrated that speaker of Pakistani English systematically adapt English sounds to the phonemic system of their native languages; for instance, /θ/ becomes [t̪], vowel diphthongs are neutralized, and /r/ is pronounced in all positions. Such processes indicate nativisation, in which English has developed under a certain sociolinguistic context to constitute an ongoing, self-sustaining local variety.

This study offers an array of significant advancements in the three areas of computational linguistics, phonology, and language pedagogy. The research provides a necessary series of connections between linguistic description, technological application, and educational practice, lending to the overall academic and practical utility aspects of research on Pakistani English.

REFERENCES

Abbas, S., Umar, A., Bilal, H. A., Raza, M. U., & Safdar, A. (2020). Acoustic analysis of consonants of Pakistani English. Ilkogretim Online - Elementary Education Online, 19(4), 4405–4429. https://doi.org/10.17051/ilkonline.2020.04.764846

Malik, M. Z. A., Shoukat, B., & Kharal, A. A. (2022). Acoustic analysis of the socio-phonetic variation in the English language long vowels. Pakistan Journal of Humanities and Social Sciences, 10(1), 50–56. https://doi.org/10.52131/pjhss.2022.1001.0173

Shahid, A., Mirza, H. M. M., Raza, M. U., & Bilal, H. A. (2024). A survey of phonological characteristics of Pakistani English. Migration Letters, 21(S10), 957–971.

Bilal, H. A., Faiz, R., Asghar, I., & Abbas, S. (2025). Mapping segmental variations in Pakistani English: A sociophonetic study. 3L: The Southeast Asian Journal of English Language Studies, 31(2), 194–209. https://doi.org/10.17576/3L-2025-3102-14

Yaqub, M. (2025). Acoustic analysis of vowel variation in Pakistani English: 13 varieties 13 cities. Journal of Applied Linguistics and TESOL, 39(1), 15–33.

Tufail, S., & Bin Sadiq, A. H. (2025). Sociophonetic analysis of English triphthongs in Pakistani English: Pronunciation patterns and phonological implications. Journal of Arts and Linguistics Studies, 3(1), 969–1002. https://doi.org/10.71281/jals.v3i1.257