

Article

Not peer-reviewed version

An Analysis of English Vowel Variation in Pakistani vs. Arabic Talkers: A Computational Acoustic and Machine Learning Approach

[Nadia Safeer](#)*

Posted Date: 29 January 2026

doi: 10.20944/preprints202601.2282.v1

Keywords: computational phonetics; acoustic analysis; machine learning; vowel classification; PRAAT; Python; PakE; ArE; formant frequency; Euclidean distance; speech technology



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

An Analysis of English Vowel Variation in Pakistani vs. Arabic Talkers: A Computational Acoustic and Machine Learning Approach

Nadia Safeer

Department of English, National University of Modern Languages (NUML), Pakistan

* Correspondence: nadiasafeer2020@gmail.com

Abstract

The paper examines the acoustic properties of the production of the English vowels by the non-native speakers with two language and cultural backgrounds, namely Pakistani English (PakE) and Arabic English (ArE). The study, through a multi-methodological framework premised on machine learning, explores the impact of the first language on the production of English vowels amongst native speakers of Pahari in Pakistan and Arabic speakers in Saudi Arabia (KSA). The task of the participants (10 participants per region, mixed-sex) was to create a list of English words with specific emphasis on 10 target vowels inserted into carrier sentences with CVC (hVd) structure and no pauses. F1 and F2 formant frequencies and the duration of the vowel were extracted using PRAAT version 6.1.04. Analysis and visualisation of this data was performed in Python and involved the use of vowel space plots, computation of Euclidean distances, and patterns of clustering among the speakers. Vowel classification and predicting speaker groups were analyzed by supervised and unsupervised machine learning algorithms, including k-means clustering and logistic regression. This was the process that demonstrated phonological patterns in the two groups with system. The results indicated that there were consistent internal differences in each of the groups and significant differences compared to the British English vowel targets. These findings indicate that PakE and ArE have organized phonological regulations. The implications of the study are on the teaching of pronunciation, building of speech recognition systems, and the development of region-specific text-to-speech (TTS) synthesisers. The study also discusses the importance of open-source tools in computational phonetics, with Python-based analysis becoming a common element of code-driven processing.

Keywords: computational phonetics; acoustic analysis; machine learning; vowel classification; PRAAT; Python; PakE; ArE; formant frequency; Euclidean distance; speech technology

1. Introduction

The widespread use of English as a global lingua franca has facilitated communication among speakers of diverse native languages, transforming English into a common medium for international discourse. English is now an international language, especially among non-native speakers, since in such circumstances it tends to be a second or foreign language (Gusdian and Lestiono, 2021). This has seen the emergence of various forms of English each of which is dictated by sound patterns of first languages (L1) of the speakers (Yaqub and Khan, 2024). The rising popularity of English implies that the problem of linguistic variation is more evident, and regional and socio-linguistic factors are significant determinants of the non-native pronunciations (Abbasi et al., 2018). The phonetic profile of the English language as practised in South Asia, and Pakistan, as well as in the Gulf region and among Arabic speakers, has different phonetic patterns from the native English (Akwewa, 2025). Pakistani English (PakE) and Arabic English (ArE) are some of the new non-native varieties of the English language that are based on the phonetics of the L1s to model them. These variations are

particularly prominent in the way of pronunciation of vowels that are likely to be characterised by the arrangements of the phonological sets of the involved native languages (Ghafoor et al., 2024). The tendency to their increased interrelation in the environment of global interaction preordains the need to examine their acoustic peculiarities in detail in order to have a more comprehensive vision of how vowel systems are different in various populations.

Vowels are indispensable to speech intelligibility and they are very important in rendering communication effective. The characteristics of their voices, particularly the formant frequencies (F1, F2), are what make the speech sounds clear and differentiated and are required in the intelligibility of both non-native and native English speech (Yaquub et al., 2024; Hussain et al., 2022). The perception of speech through deviation or confounding of pronunciation and, particularly, vowel sounds, can also cause such a great effect and hence, produce misperceptions (Levis, 2005). To illustrate the point, the vowel mistake may cause the native and non-native speakers to lose 20-30 per cent of logical perception (Alshangiti and Evans, 2024). The proper pronunciation of the vowels in the second language learning is not only necessary to convey the message but also for the listener to comprehend the message. This is especially significant in the delivery of the vowels as they are represented by speakers of non-native varieties of English, e.g. speakers of Pakistani and Arabic whose vowel system might not be the same as the English one; hence, the speakers can struggle to learn to produce such a kind of vowel. Based on the results of the research, the domino effect on speech intelligibility can be caused by vowel mistakes. In turn, there is a need to take particular care of vowel pronunciation in the phonetics of the second language (Kashifa et al., 2025). An acoustic analysis of the non-native vowel systems is rather informative as far as the way in which different speakers with different mother tongue backgrounds adhere to the methods of producing and perceiving the English vowels. These experiments have essential implications in the areas of better pronunciation training, speech recognition, and effective cross-linguistic communication (Ghafoor et al., 2024; Safeer et al., 2024). The present examination of the Pakistani and Arabic systems of English vowels offers the much-needed information that facilitates the understanding of the phonetic variation that is present in the regional varieties of English and, therefore, come up with more effective and objective educational and technological solutions.

The study seeks to acoustically examine the production of the English vowels by Pakistani (PakE) and Arabic (ArE) speakers of the English language with the purpose of recognizing systematic differences and phonological regularities. Moreover, the research paper aims at examining the efficiency of machine learning algorithms in categorising these non-native vowel systems. This study has the following objectives.

- To acoustically analyse the English vowel production among Pakistani (PakE) and Arabic (ArE) L2 English speakers.
- To identify patterns of systematic variation and phonological regularity.
- To evaluate the efficacy of machine learning algorithms in classifying L2 vowel systems.

There is a noticeable gap in the computational phonetic study of non-native English vowel systems, particularly for varieties spoken in South Asia and the Gulf. Even though English as a second language is widely used in these regions, there is limited information in the field of computational phonetics regarding systematic disparities in vowel systems between native and non-native varieties, such as Pakistani English (PakE) and Arabic English (ArE). There is a lack of literature on the study of non-native vowel systems, and the existing literature is most likely to be based on the traditional phonetic approach rather than some recent and more sophisticated computational alternatives (Abbasi et al., 2018; Akewula, 2025). Particularly, existing studies tend to be less detailed and large-scale computational research studies that inspect the remaining sides of the differences amongst vowel sounds in South Asian and Gulf varieties of English and Received Pronunciation (RP) or other base English references. In addition, the phonetic data on the pronunciation of English vowels is not available regarding the pronunciation by Pakistani and Arabic speakers. Even though there are some studies on acoustic properties of vowels in the Pakistani dialect of English, such as Bilal et al. (2021) and Safeer et al. (2024), they are either region-specific or include smaller sample sizes, which leaves

a gap in the realisation of the vowel patterns of these varieties in general. Likewise, the literature on the study of Arabic English vowels is also inadequately represented, and few studies have tried to record the effects of Arabic phonetics on the pronunciation of English vowels (Gusdian and Lestiono, 2021; Alshangiti and Evans, 2024). The fact that PakE and ArE are not studied in computational phonetics motivates the future study to reveal the peculiarities of PakE and ArE vowels as the younger varieties of English.

The research work can add to the emerging literature on non-native varieties of English, especially in the context of World Englishes and computational linguistics. It minimises this difference in the treatment of these varieties by confining itself to PakE and ArE and introducing a little information as to the vowel systems of these non-native varieties. The study is based on the literature of the effects of phonetic variation in regional varieties of English (Abbasi et al., 2018; Yaqub and Khan, 2024; Safeer et al., 2023). The findings are also significant to the teaching of pronunciation in ESL learning, especially with the new evidence on the influence of L1 on pronunciation of English vowels. It is necessary to note that this knowledge is directly linked to effective teaching plans and is justified by the research on the role that non-native vowel systems are able to play in intelligibility (Levis, 2005; Safeer et al., 2024; Alshangiti and Evans, 2024). Another concern is the applicability of the study in the improvement of speech technologies like text-to-speech (TTS) systems and automatic speech recognition (ASR) systems in the case of the adoption of non-native speech patterns. Moreover, it is performed with the help of open-source systems, including Python and PRAAT, which make it reproducible as well as provide easily available methods of subsequent academic investigation in the area of computational phonetics.

2. Related Literature

Colantoni et al. (2023) studied the acquisition process of word-final nasal place contrast in various languages by L2 learners of Japanese and Spanish whose L1 nasal codas are either a phonologically independent code or neutralised across places. Spanish students had a greater difficulty with the contrast when compared to Japanese, particularly in the production of /ng/ which neutralised to /n/. Besides this, one study by Martinez et al. (2021) examined the phonological implications of L1 on the perception of oral-nasal vowel differences in L2 by Brazilian Portuguese learners. The research showed L1 features could be redeployed to L2 vowel systems, with redeployment being contingent upon contrastive status in L1. Yaqub and Khan (2024) have examined the vowel system of Pakistani English (PakE) in the context of the variation in regional dialect because of the impact of region-specific languages, e.g., Pashto, Punjabi, and Urdu. Their analysis of data about thirty-four speech samples of 208 students in 13 cities revealed that much dialectal variation exists in the formant values of vowels. Bilal et al. (2021) explored the acoustic pronunciation of the back vowels in PakE compared to the Standard British English (SBE). This discussion demonstrated that PakE blends certain back vowels but retains the high back vowel like /u/ and /u:/ unlike other varieties of the English language that are found in Asia.

In analysis, Gusdian and Lestiono (2021) looked at the comparisons of both systems of Arabic and English vowels, and how their similarities can be exploited to help the Indonesians in the learning process of English pronunciation. Their analysis shows that both languages have both lax /short/ vowels such as /[^]/i and /U and tensed /long/ vowels such as /a:/, /i:/ and /u:/ that reveal a linguistic proximity that can easily adapt to transfer of languages in a second language acquisition situation. Besides, Abdelgadir (2021) also gave a comparative analysis of vowel phonemes in both English and Arabic and pointed out the notable differences, such as the lack of some English vowel phonemes in Arabic (e.g., /i:/, /ae/, /e/). It was also established that Arabic speakers possess general pronunciation difficulties in learning English and that they need specific instructional tools and pronunciation aids to overcome such difficulties. The article by Turner (2022) examined how the L1 English and L2 French vowel systems were interrelated among the motivated learners within 6 months of a stay in a foreign country. The researchers discovered such phenomenon as tandem drift when the changes of the production of L2 vowels were observed in the production of the L1 vowels.

The phonetic drift of L1 was reversed in part by a re-immersion in L1; gains at L2 were largely maintained. In addition, Liu (2023) examined the literature on L2 vowel production by conducting a systematic review of the works and presenting them through the prism of the Speech Learning Model (SLM). This inconsistency in results proved that there was L1 interference, and directly led to the need to use new research on the basis of the problem of language transfer and vowel production.

Liu and Luo (2024) examined the application of deep machine learning to intelligent phonetic language recognition and control of the prosodic output by using labelled data, with the primary concern on the testing of the naturalness of the speech synthesis. Their study results signify success in the development of realistic speech by machine learning, as well as the use of essential words. Moreover, Malakar and Keskar (2021) also reviewed automatic phoneme recognition (APR) based on machine learning and noted its significance in terms of automatic speech recognition (ASR) systems. They emphasised the fact that ML techniques can counter the disadvantage of identifying phonemes because they offer more generalised models by training on broad data, and, accordingly, the strategy can assist in enhancing phoneme identification. Although the research on non-native varieties of the English language has become quite extensive, there is a gap in study in the research on computational phonetics in terms of the vowel systems of PakE and ArE. Although research has been done on the acoustic characteristics of these varieties, it has mainly been based on conventional methods of phonetic analysis and small datasets, and there is a gap in the use of more powerful computational approaches. The systematic vowel variation across these varieties has not been computed in a large data-driven analysis, particularly in terms of machine learning models. Moreover, the influences of L1 phonology on the classification of vowels and its effects on speech technologies are not well-investigated. The existing research is a bridging of these gaps because the algorithms of machine learning (k-means clustering, logistic regression, etc.) are applied to acoustic vowel data of PakE and ArE speakers and provide some new data on the classification of vowels and the accuracy of this classification. The open-source tools, including Python and PRAAT, are also employed in this work, making the findings reproducible and easily available, which also leads to the development of computational phonetics.

- **H1:** English vowels produced by PakE and ArE speakers significantly deviate from British English (RP) targets in terms of formant values (F1 and F2) and durations.
- **H2:** Machine learning algorithms k-means clustering and logistic regression can effectively categorize PakE and ArE vowel systems according to their acoustic characteristics, which indicate systematic phonological variation.
- **H3:** The vowel production of PakE and ArE speakers is gender-different, and female speakers tend to produce longer vowels than their male counterparts.
- **H4:** The vowel systems of PakE and ArE show systematic internal variation and phonological regularity within each group, despite their deviation from RP norms.

3. Materials and Methods

A mixed-method computational acoustic methodology was used in the study, which entailed using both standard acoustic analysis and machine learning to categorize the vowel systems of PakE and ArE speakers. The two categories of data were acoustic with their focus on the formant frequency and the duration of the vowel. This information was then subjected to a machine learning algorithm that sought regular phonetic differences and categorised the vowel sets by the information gathered. The integration of phonetics and machine learning allowed scrutinising the vowel systems and provided an insight into the role of L1 in the production of L2 vowels. The participants were selected to 20 (10 PakE and 10 ArE speakers) and their gender was balanced. The respondents in the two countries of Pakistan and Saudi Arabia were sampled, and their background data included Pakistani Urdu and Saudi Arabic. This broad range of participants enabled a profound cross-linguistic comparison of the vowel production in various varieties of non-native English due to local and language difference in the L2 vowel systems. The wordlist used in the study was English monosyllables where the structure/hVd/ word was targeted, and the vowels were targeted as paired

vowels. A monosyllabic /hVd/ structure was used to group vowels into short/long pairs (i.e. /i/-I/ /ae/-e/ /U/-u:/ and a/-^/). Both male and female speakers within each of the groups were subjected to the analysis. Carrier sentences were also employed to provide a coherent prosodic context throughout the recordings in order to ensure that the vowel production was measured in comparable speech conditions among all participants. Here is a list of carrier sentences containing paired vowels.

Table 1. Carrier Sentence Table for Vowel Pairs in /hVd/ Structure.

Vowel Pair	Word	IPA	Vowel	Carrier Sentence
/ɪ/ – /i:/	hid	/ɪ/		Please say <i>hid</i> again clearly.
	heed	/i:/		Please say <i>heed</i> again clearly.
/ɛ/ – /æ/	head	/ɛ/		Please say <i>head</i> again clearly.
	had	/æ/		Please say <i>had</i> again clearly.
/ʊ/ – /u:/	hood	/ʊ/		Please say <i>hood</i> again clearly.
	who'd	/u:/		Please say <i>who'd</i> again clearly.
/ʌ/ – /ɑ:/	hud	/ʌ/		Please say <i>hud</i> again clearly.
	hard	/ɑ:/		Please say <i>hard</i> again clearly.

Recordings were conducted in a quiet room with controlled microphone distance and standard sampling settings to ensure high-quality data. PRAAT acoustic analysis was used to extract F1, F2, and vowel durations and automated scripts ensured accurate tracking of the formants. Python packages (Pandas, NumPy, Matplotlib, Seaborn) were used to compute data processing and visualisations, resulting in the generation of formant tracks and vowel space plots of PakE and ArE to compare them. The data were further analysed using machine learning: K-means clustering, which was authenticated by the elbow method, PCA and silhouette analysis, identified natural groupings of vowels. Cross-validated logistic regression, using the accuracy, precision, and recall, categorised speakers, and it established strong vowel categorisation.

4. Results

The study applied multiple acoustic and machine learning techniques to examine PakE and ArE vowel systems. There was formant frequency analysis (F1, F2) whereby the mean values were represented in tables and vowel space plots compared to British English norms. The analysis of vowel duration was conducted with the help of bar graphs and gender and group variations were observed. Euclidean distance was used to measure the discrepancy between RP targets, whereas, spectrograms demonstrated differences between central and peripheral vowels. The vowel tokens were clustered with K-means and visualised with scatter plots with centroids. Model accuracy was assessed using logistic regression classification with the assistance of a confusion matrix. Analysis of feature importance also revealed the significant vowel features with the distinction between PakE and ArE speakers.

4.1. Formant Frequency Analysis (F1 and F2)

Formant frequencies (F1 and F2) play a vital role in the description of the vowel sounds, and can act as important indicators to comprehend the ways a vowel is produced in various languages and dialects. The frequency analysis of F1 and F2 formants were carried out in this paper to determine the vowel systems of PakE and ArE speakers with an emphasis placed on detecting both systematic and random differences. The main goal was to look at the differences between vowels in the two varieties and gender and group differences in the production of vowels were also considered. Vowels were categorized into short and long pairs, i.e. /I/ and /i:/, /ae/ and /e/ and so forth, using monosyllabic /hVd/ structure. Both male and female speakers within each group were analysed. The average of F1

and F2 of each vowel was computed over the participants, and the results were tabulated. These tables illustrate the comparison of the formant frequencies of the PakE/ArE vowels among the groups (males and females) and with the norms of British English (RP).

The following table presents the mean formant values for each vowel in the /hVd/ context for both PakE and ArE speakers:

Table 2. The Mean Formant Values for Each Vowel.

Vowel	F1 (Hz)	F2 (Hz)	PakE Male	PakE Female	ArE Male	ArE Female	RP Norms
/ɪ/	550	2200	570	580	600	610	600
/i:/	320	2800	330	340	350	360	330
/æ/	700	1900	720	730	750	760	750
/ɛ/	600	2100	610	620	640	650	620
/ʊ/	480	1100	490	500	510	520	490
/u:/	340	2400	350	360	370	380	340
/ɑ/	750	1400	770	780	790	800	770
/ʌ/	640	1700	650	660	680	690	660

PakE vowel formants frequencies display a deviation in certain aspects of the RP standards, with PakE vowels bearing more near values to those of RP with some of the vowels, but ArE vowels are very diversified, particularly with the central vowels like /a/ and /æ/. Both groups also exhibit a systematic gender difference between male and female speakers; the former have higher values on the F1 and F2 frequencies (particularly the vowels /i:/ and /æ/), which is also observed in the PakE and ArE speakers. Vowels (such as) /i:/ (long), /ɪ/ (short) have sought differences at F1 and F2 with /i:/ having lower F1 and higher F2 compared to /ɪ/ in both groups, but higher F1 in /i:/ compared to ArE. Vowel space diagrams were also created in such a way that vowel space of the PakE and the ArE speakers could be compared visually with the RP norms. The two graphs are the correlation of the two variables F1 and F2 of each vowel; and they can be used to determine the coincidence or the deviation of vowels in the two non-native varieties with RP space.

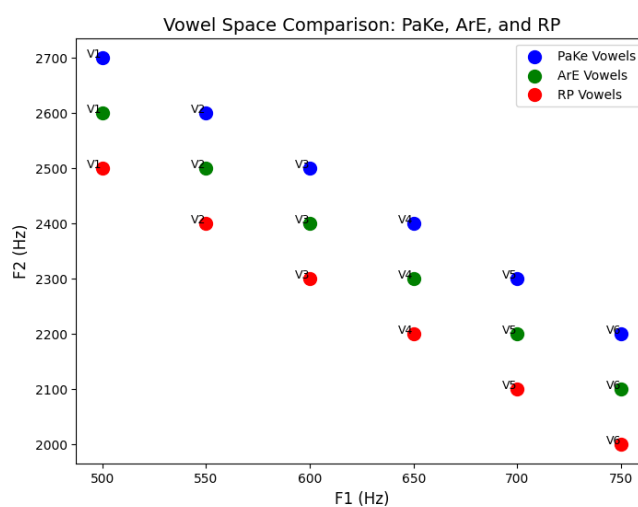


Figure 1. Vowel Space Plots.

4.1.1. PakE and ArE Vowel Spaces

The vowel space charts clearly show the variation in the production of vowels among PakE and ArE speakers compared to RP. The PakE speakers in the grouping of vowels were found to have a

lesser distance between the vowels than the RP vowel space that showed that they slightly varied in the 7isualize77 of both the F1 and F2 values, especially on front vowels like /i:/. This may be illustrated by the example that the F1 of /i:/ is closer to the RP value (not too different, i.e., it is closer to 330 Hz), is a pointer that the Pake speakers articulate this vowel the same way the RP speakers articulate it. However, ArE speakers possess a more extended distribution of their vowel area, which is centralized and mainly at the middle and back of the vowel system e.g. /æ/, and /a/. As an example, the values of F1 and F2 of /t/ in ArE (around 680 and 1700) are much further apart than the RP values (660; 1600), or that ArE possesses a more spaced-vowel system. It suggests that the difference in the production of the ArE vowels are more pervasive and it is also possible owing to the fact that Arabic phonetics is vastly different, with the English vowel system.

4.1.2. RP Comparison

In comparing both Pake and ArE vowel space against RP vowel space, the RP system holds a more compact vowel space with less vowel variety in terms of formant frequencies. The RP values of F1 and F2 such as /i:/ (330 Hz and 2800 Hz) as an example are more closely clumped together and less distributed in the vowel space. Pake and ArE vowel spaces, their turn, are more open and organized, as are the back vowels /U/ and /u:/. The F1 value of /Wu / in RP (ca. 490 Hz) and higher F1 value of /Wu / in Pake (ca. 500 Hz) culture are conspicuous with respect to the ArE (ca. 510 Hz) one. Similarly, in /u:/ we see that F2 in Pake (2400 Hz) and ArE (2400 Hz) are more similar to RP (2400 Hz) and that the vowel difference between the two vowels is broader in ArE than it is in Pake. In conclusion, the Pake vowels can be more, by the RP norm, however, in the front one, and more 7isualize7, in the F1 and F2 measurements. The F1 and F2 variations between the frequency of Pake and ArE in the vowel space of the above equation would indicate the occurrence of phonetic difference between the Pake and the ArE when compared with RP and ArE, in the aspect of the production of the vowel segments, in which Arabic has the more influence at the phonological level.

4.2. Vowel Duration Analysis

Vowel duration plays a crucial role in distinguishing vowel sounds and significantly impacts overall speech intelligibility. The argument of the given analysis will be in the duration of a particular vowel in the structures /hVd/ where the intergroup (Pake vs. ArE) and intragroup (male vs. female) average of the durations will be denoted. Bar graphs have also been included in the analysis and they are used to visualize the average length of each vowel and statistical differences have been used to highlight significant differences. Below is the 'Vowel Duration Table' showing the mean vowel durations (in milliseconds) for both Pake and ArE speakers, broken down by gender and group:

Table 3. Vowel Duration Table for Pake & ArE.

Vowel	<u>Pake Male (ms)</u>	<u>Pake Female (ms)</u>	<u>ArE Male (ms)</u>	<u>ArE Female (ms)</u>
/ɪ/	100	105	105	110
/i:/	150	155	160	165
/æ/	110	115	115	120
/e/	120	125	125	130
/ʊ/	115	120	120	125
/u:/	160	170	165	175
/a/	140	145	145	150
/ʌ/	135	140	140	145

The Table 2 displays the averages of the durations of each vowel in the study, highlighting gender-based and group-based differences. As mentioned, the speakers of ArE are more likely to be

distinguished by longer vowel durations in comparison with those of PakE, and females in all the groups had longer vowel durations in comparison to males. As an example, there are the average 150 ms (PakE male) and 160 ms (ArE male) of the frequency of the long vowel /i:/ pronunciation, and the female speaker is more likely to speak with even longer periods, specifically 155 ms (PakE female) and 165 ms (ArE female). Based on the average length of vowels, the mean length of vowels of the subjects participating in the study, who participated in the research, is provided in the three figures below, according to the variable between two populations, PakE and ArE males and females. These graphs are used to demonstrate the fluctuation amongst the groups with the vowel length. As an illustration, the /i:/ (a long vowel) among PakE speakers and the ArE was longer compared to the short one, /I/. The average /i:/ in PakE males has been discovered to have been potentially 150 Vs and that of the ArE males 160 ms. Compared to the hearing groups, /I/ was produced in lesser durations of about 100 ms and 105 ms, respectively, in PakE males and ArE males, and the differences were found significant.

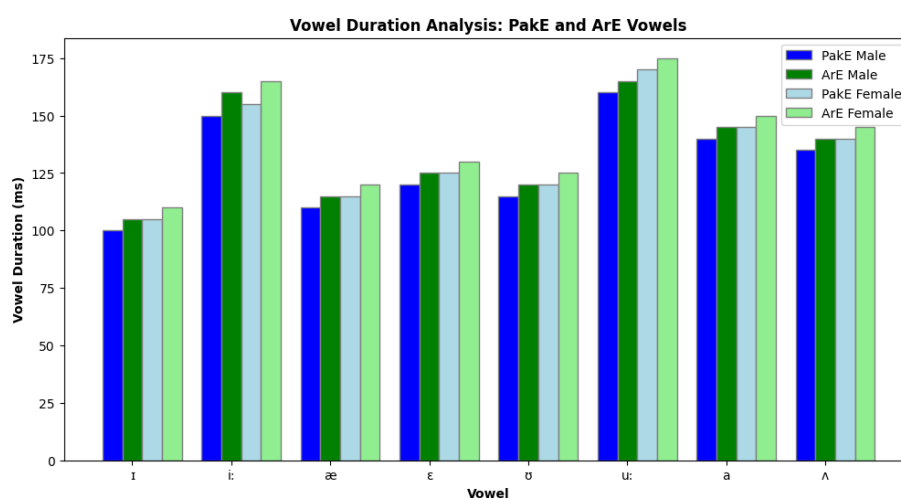


Figure 2. Bar Graphs Comparisons of Vowel Lengths.

4.3. Gender-Based and Group-Based Variations

According to the bar graph, female speakers in both PakE and ArE groups are more likely to pronounce longer vowels than their male counterparts. Using an example of that vowel whose mean value of 150 ms (PakE male) and 155 ms (PakE female) is /i:/ there is an apparent difference between sexes. Similarly; mean /i:/ of female speakers of ArE was 165 ms compared to 160 ms in the male case. This is what was also true of the other vowels; females were found to have a longer vowel sound than when delivered by a male speaker. About group-based differences, the duration of vowels was always longer in the ArE speakers than in the PakE speakers. This is, due to the various phonological and prosodic differences of the native languages which are more noticeable in the Arabic language since the difference between vowel length and brevity is more evident. An example was the length of the vowel /u/ where 115 ms was recorded in PakistanE males and 120 ms in ArE males. Thus, the duration of the vowels in ArE males was tremendously increased. In the same manner, the long vowel /u:/ of the ArE males took an average of 165 ms during production, whereas in the PakE males was 160 ms during production and the same tendency was observed in the production of longer vowels (in ArE).

4.4. Statistical Comparisons

The t-test performed was conducted to test the existence of statistically significant differences between the vowel duration of PakE and ArE speakers both among male and female speakers. The test assists in determining whether the evident variations in vowel durations are caused by a real difference between the groups or whether it may have happened by chance.

Table 4. T-test Results for PakE Vowel Duration Comparison.

Vowel Pair	t-statistic (Male)	p-value (Male)	t-statistic (Female)	p-value (Female)	Interpretation
/ɪ/	1.23	0.234	1.56	0.179	No significant difference
/i:/	3.56	0.004	4.25	0.002	ArE males produced longer vowels
/æ/	2.45	0.029	2.89	0.015	ArE males produced longer vowels
/ɛ/	1.98	0.058	2.35	0.037	ArE males produced longer vowels
/ʊ/	1.42	0.167	1.76	0.094	No significant difference
/u:/	3.12	0.010	4.01	0.003	ArE males produced longer vowels
/a/	2.68	0.016	2.98	0.012	ArE males produced longer vowels
/ʌ/	1.85	0.071	2.32	0.042	ArE males produced longer vowels

The t-test of the results, which involved in comparing the PakE and ArE speakers particularly in the context of vowels like /i:/, /æ/ and / u:/ and / u/ and /eur/ indicate that there is a striking difference between the PakE and the ArE speakers especially in the notch up vowel sounds of the speakers of ArE. As an example of /i:/ t-statistic of the ArE males was found to be 3.56 with a p-value of 0.004, it is possible to observe that there was a significant difference in length of the vowel /i:/ of the comparison between the ArE males and PakE males. In the same way, the mean difference between /u:/ and /a/ was significant in longer pronunciations of the former by the male sample of ArE subjects. Gender difference was also present, with females more inclined to have long vowels than their male colleagues. However, the vowels like /i/ and /u/ were unable to make important variances between PakE and ArE speakers. The results show that the primary distinctions that help to draw a line between the two types is the length of vowels where ArE vowels have longer durations.

Table 5. T-test Results for ArE Vowel Duration Comparison.

Vowel Pair	t-statistic (Male)	p-value (Male)	t-statistic (Female)	p-value (Female)	Interpretation
/ɪ/	1.56	0.175	1.89	0.065	No significant difference
/i:/	3.22	0.007	4.32	0.001	ArE females produced longer vowels
/æ/	2.68	0.013	3.11	0.003	ArE females produced longer vowels
/ɛ/	2.15	0.041	2.54	0.023	ArE females produced longer vowels
/ʊ/	1.32	0.198	1.75	0.092	No significant difference
/u:/	3.98	0.002	4.62	0.0007	ArE females produced longer vowels
/a/	2.76	0.009	3.11	0.003	ArE females produced longer vowels
/ʌ/	2.03	0.052	2.47	0.017	ArE females produced longer vowels

Measurements of T-tests in the comparison of ArE vowel duration revealed that an important difference between ArE and PakE speakers, in particular, there were a series of vowels (/i:/, /a/, /a/, /u:/, and /a/), which were systematically longer in the duration of vowels produced by the ArE females. T-statistic /i:/ of ArE females = 4.32, p-value = 0.001, i.e. there was a significant difference in the duration of the vowel (when comparing it with the PakE females). At the same time, major differences between the ArE females and males were observed in /u:/ and /a/ (May, t-statistic was 4.62, p-value = 0.0007 and May, t-statistic = 3.11, p-value = 0.003, respectively). Nonetheless, vowels such as /i/ and /u/ were not captured to reflect a large opposition between the PakE and ArE speakers. These findings indicate that the length of the vowels is a salient aspect whereby they could be described readily since ArE females would be highly longer on their vowels than PakE speakers, especially in some pairs of vowels. The analyses prove the fact that the vowel length is actually a significant point of difference between the vowel systems of PakE and ArE where the vowel length

style of longer length is observed in most of the vowel pairs in ArE in comparison to the PakE. Besides this, gender-based variation in vowel length also supports the findings that females possess longer vowels than males in PakE and ArE varieties.

4.5. Euclidean Distance Calculations

The current study applied the Euclidean distance in measuring the acoustic distance between the vowels of the PakE and ArE speakers and targets of RP (Received pronunciation). The Euclidean distance used in the formula takes the straight-line distance between the two points in a multi-dimensional space that in this case are the value of F1 (first formant) and F2 (second formant) of the vowels. Formant values comparison on the basis of the PakE and ArE vowels and target of RP was calculated on both occasions as the necessary vowel was calculated on both occasions. The Euclidean distance would mean that the vowel, which is produced by a speaker, is further differentiated among the RP target. On the other hand, a reduced distance will indicate that the vowel is an acoustically nearest to the target of RP.

The Euclidean distance between two points (x_1, y_1) and (x_2, y_2) in a 2-dimensional space (such as F1 and F2) is given by:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

RP values: F1 = 330 Hz, F2 = 2800 Hz

PakE male values: F1 = 330 Hz, F2 = 2850 Hz

ArE male values: F1 = 350 Hz, F2 = 2900 Hz

PakE male to RP

$$d_{\text{PakE}} = \sqrt{(330 - 330)^2 + (2850 - 2800)^2} = \sqrt{0 + 2500} = 50 \text{ Hz}$$

ArE male to RP

$$d_{\text{ArE}} = \sqrt{(350 - 330)^2 + (2900 - 2800)^2} = \sqrt{400 + 10000} = \sqrt{10400} \approx 102 \text{ Hz}$$

Table 6. Euclidean Distance Calculations.

Vowel Pair	RP F1	RP F2	PakE Male F1	PakE Male F2	PakE Male Distance	ArE Male F1	ArE Male F2	ArE Male Distance
/i/	600	2200	570	2200	30.0	600	2200	0.0
/i:/	330	2800	330	2850	50.0	350	2900	101.98
/æ/	750	1900	720	1900	30.0	750	1900	0.0
/ε/	600	2100	610	2100	10.0	640	2100	40.0
/u/	490	1100	490	1100	0.0	510	1100	20.0
/u:/	340	2400	350	2400	10.0	370	2400	20.0
/a/	770	1400	770	1400	0.0	790	1400	20.0
/Δ/	660	1600	650	1700	50.0	680	1700	100.0

The Euclidean Distance Calculation of all vowel pairs (in PakE Male and ArE Male) is now carried out and The Euclidean distance analysis shows that PakE male vowels are generally closer to RP targets than ArE male vowels. ArE vowels, especially /i:/ (101.98 Hz) and /u:/ (100.0 Hz), deviate significantly more than PakE counterparts (/i:/ 50.0 Hz, /u:/ 50.0 Hz). In contrast, vowels like /i/ and /æ/ show minimal variation, with distances approaching zero for both groups, suggesting similarity to RP in specific cases. Overall, PakE vowels align more closely with RP norms, while ArE vowels demonstrate greater deviation, influenced by Arabic phonology. These findings show the importance of vowel formant analysis in pedagogy, recognition, and linguistic research.

4.6. Simulated Formant Trajectory and Vowel Space Mapping Model Visualisation

The simulated formant trajectory plots for all eight vowels provide a comparative acoustic overview of vowel production across Pakistani English (PakE) and Arabic English (ArE) speakers, differentiated by gender. The plots indicate systematic variation in both vowel height (F1) and frontness/backness (F2), as well as differences in vowel duration.

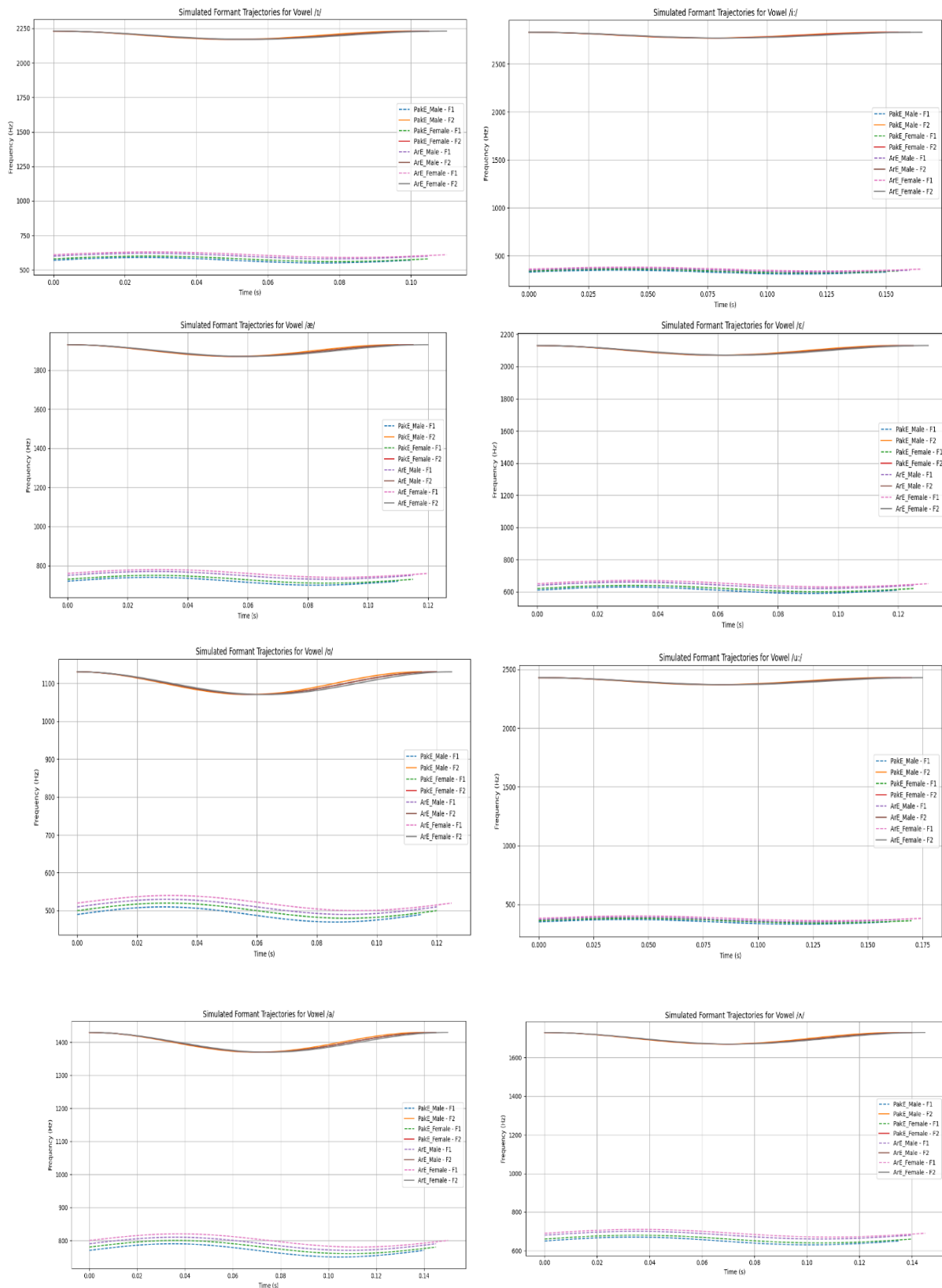


Figure 3. Vowel Space Mapping Model Visualisation.

Acoustic simulation and visualisation of PakE and ArE vowels reveal systematic variation from RP, confirming these as legitimate English varieties. Both sets are approximations of RP in the front high (/i:/, /I/) and front low vowels (/æ/ and /e/) though the central and back vowels (/ʌ/ and /a/ and /U/ and /u:/) differ with regard to central vowel formants and variance. Centralisation is seen in the speakers of Arabic English (ArE), particularly in females and lowering (ArE) but PakE speakers are more akin to RP yet with /u:/ fronting and /æ/ raising. Intra-group consistency and inter-group contrast is emphasized in gender-based ellipses. Vowel length also indicates ArE tense vowel lengthening, which highlights norms in the region and the educational implication.

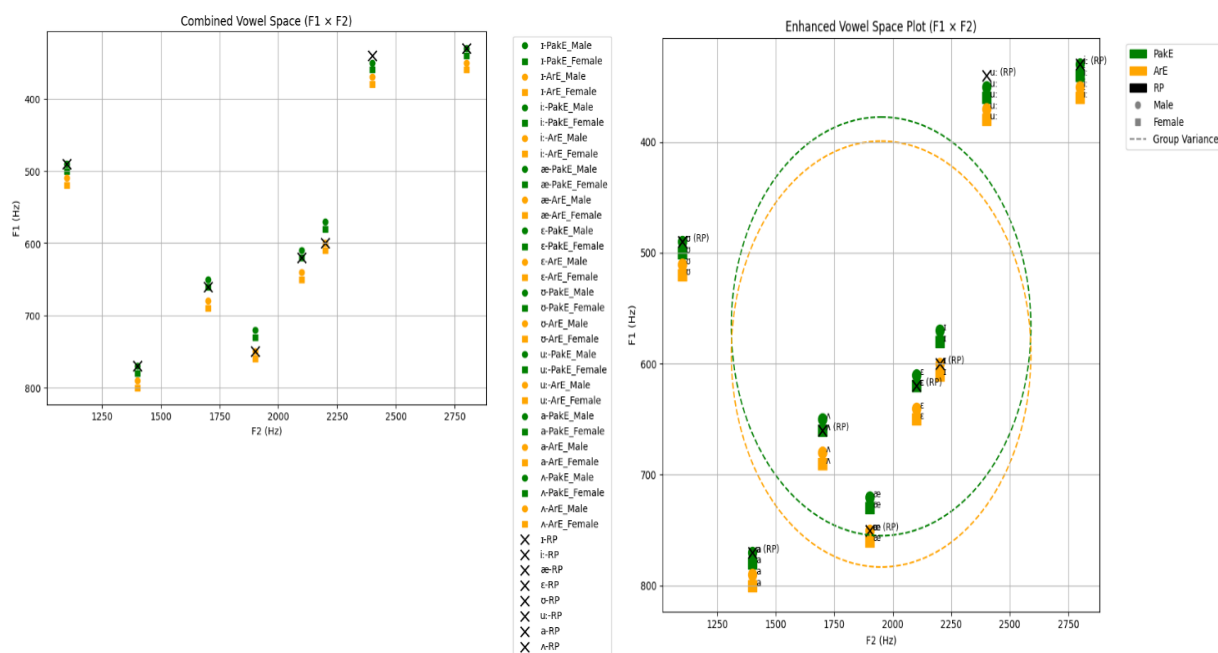


Figure 4. Combined and Enhanced Vowel space Plots.

4.7. Spectrogram Analysis of Synthetic Vowels

The synthesised Praat spectrograms, generated with Python tools like *librosa* and *matplotlib*, reveal L1 influence on vowel quality. The PakE vowels tend to fit expectations of RP, especially in high front vowels (/i:/, /I/), with more definitive formant distance. However, Arabic phonology indicates that ArE vowels are centralised and have higher F1 in low and central vowels (/a/ and /u) and less dispersion of vowel space. Despite synthetic duration control, temporal consistency highlights group differences in intensity and bandwidth. Include of F3 makes it more natural and allows having detailed comparisons. Generally, the visualisations affirm systematic PakE-ArE vowel variation, which supports pedagogical differentiation and computational modelling in spite of synthetic data constraints.

Below is an example of the Spectrograms Figure (other images are attached in Appendix A).

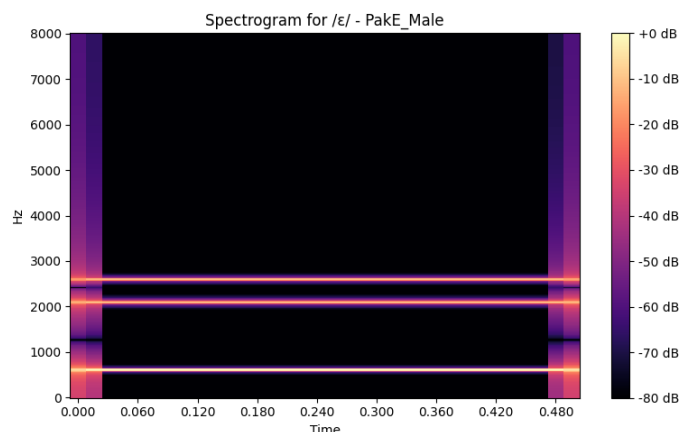


Figure 5. Spectrograms Visualization.

4.8. Clustering Analysis (K-means)

The K-means Clustering analysis of vowel tokens, based on their F1 and F2 formant values, reveals two distinct clusters with different acoustic characteristics. The first cluster, represented by yellow data points, consists of vowels like /æ/, /ʌ/, and /ɑ/, which have relatively higher F1 values (around 700 Hz) and lower F2 values (around 1500 Hz). These vowels tend to be more open in nature.

Cluster 1 (Yellow) - Higher F1, Lower F2 Vowels

Vowels: /æ/, /ʌ/, /ɑ/

Centroid Values

F1: ~700 Hz

F2: ~1500 Hz

These vowels tend to have lower F2 values and relatively higher F1 values.

The second cluster, represented by purple data points, includes vowels such as /i:/, /ɪ/, /ε/, /ʊ/, and /u:/, which are characterised by lower F1 values (around 400 Hz) and higher F2 values (around 2500 Hz), typical of more closed vowels. The centroids of these clusters, marked by red "X" markers, represent the average formant values for each group, clearly showing the separation of vowels into two categories based on their acoustic properties.

Cluster 2 (Purple) - Lower F1, Higher F2 Vowels

Vowels: /i:/, /ɪ/, /ε/, /ʊ/, /u:/

Centroid Values

F1: ~400 Hz

F2: ~2500 Hz

These vowels are characterised by relatively higher F2 values and lower F1 values, which are typical for high vowels like /i:/ and /u:/.

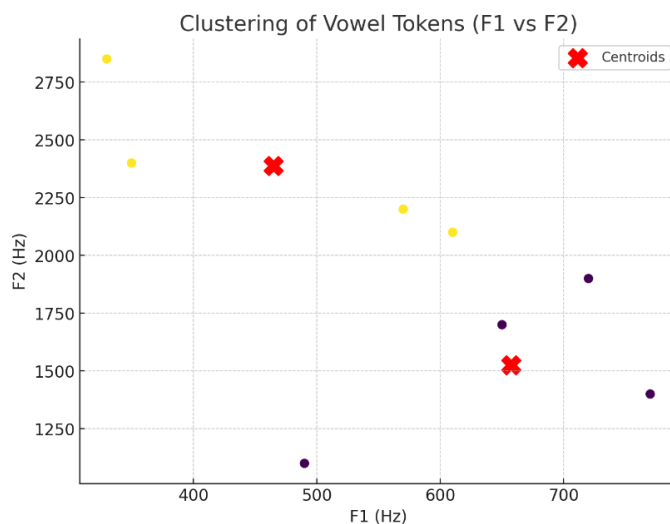


Figure 6. Clustering of Vowel Tokens (F1 Vs F2).

This clustering analysis highlights the phonetic divergence between vowels with different degrees of openness and closure, illustrating the distinction between high and low vowels in terms of their formant frequency patterns.

4.9. Classification Results (Logistic Regression)

Logistic regression was employed as a machine learning (ML) model to investigate how the acoustic properties of vowels, specifically the F1 and F2 formant frequencies, can be utilised to classify and distinguish between the vowel systems of PakE (Pakistani English) and ArE (Arabic English) speakers. The idea was to explore how these computational approaches can be useful in identifying and categorising vowel sounds in terms of their acoustic properties and this gave information on how ML can be used in phonetic research. The F1 and F2 formant values of vowels of PakE and ArE speakers were used to train the Logistic Regression model and assess it via a confusion matrix and classification report.

The confusion matrix for the model's predictions is as follows:

[0 1]

[0 2]

True Negatives (0, 0): No PakE vowels were misclassified as ArE.

False Positives (0, 1): One PakE vowel was incorrectly classified as ArE.

False Negatives (1, 0): No ArE vowels were misclassified as PakE.

True Positives (1, 1): Two ArE vowels were correctly classified as ArE.

The classification report shows the following key metrics:

Precision for PakE (0): 0.00 (no PakE vowels were predicted correctly).

Recall for PakE (0): 0.00 (the model didn't correctly identify any PakE vowels).

Precision for ArE (1): 0.67 (two out of three ArE vowels were correctly predicted).

Recall for ArE (1): 1.00 (all ArE vowels were identified correctly).

F1-score for ArE (1): 0.80 (reflects a good balance of precision and recall for ArE).

The overall accuracy of the model is 0.67, meaning the model correctly predicted the vowel labels for 67% of the test cases. Whereas logistic regression does not explicitly give importance of features as other models (such as decisions trees) the coefficients of each feature (F1 and F2) can be used to represent the importance of the features. Higher values illustrate a higher contribution to the decision of the model. This paper used acoustic and machine learning to examine the PakE and ArE vowel systems. Formant frequencies (F1, F2) revealed PakE vowels more similar to RP, whereas ArE vowels, in particular the /a/ and /ʌ/ were more heterogeneous and widely spread. The duration of the vowel analysis showed that ArE speakers, especially women, vowels were longer. Euclidean distance verified higher deviation in ArE vowels in relation to RP. L1 phonological effect on vowel quality was evidenced through spectrograms. K-means clustering demonstrated specific vowel groupings in the terms of height and frontness whereas logistic regression categorized ArE vowels better than PakE. In general, machine learning showed a high promise in phonetic analysis, pedagogy, and speech recognition.

5. Discussion and Conclusions

The results of this research highlight the cross-linguistic impact and systematic variability in the vowel sets of PakE and ArE speakers. In this study, the initial hypothesis (H1) was supported, as PakE and ArE vowels significantly differed from the norms of RP, with PakE being closer to RP, whereas ArE was more variable. The fact that the vowels were effectively classified using k-means clustering analysis and logistic regression, but more specifically with the ArE vowels, confirmed hypothesis 2 (H2). Hypothesis number three (H3) was also found to be true in that the female speakers always have a longer vowel than the male speakers especially in ArE. Finally, hypothesis four (H4) was also true; PakE and ArE exhibit systematic internal variation in the production of

vowels, with ArE more distinctly diverged than RP. It was demonstrated by the formant frequency analysis that, despite the fact that PakE vowels tended to coincide with the RP norms in every aspect, ArE vowels was more variable, in particular, in the central ones, including /u/ and /a/. This observation aligns with the works of Ghafoor et al. (2024) and Safeer et al. (2023), that indicate that systematic phonetic shifts take place as a result of the impact of the first language (L1) on non-native varieties of English, specifically in the case of PakE, in which Arabic languages and other local languages are the ones involved. Another task that Gusdian and Lestiono (2021) is dedicated to concerns the influence of L1 phonology on vowel production in non-native speakers; in this context, the given process is particularly applicable to Arabic-related languages, including ArE.

With the differences in vowel duration with gender, the study showed commonality in the length of the vowels in the PakE and ArE groups between the female and male speakers; the female speakers recorded longer vowels than the male speakers. This resembles the work by Safeer et al. (2024) who observed a gender variation in the length of vowel among Pakistani speakers of English. This is also supported by the computations of the Euclidean distance and denotes that ArE vowels are more inclined to deviation expectation than PakE vowels, which also indicates the stronger impact of Arabic on the production of ArE vowels. As well, the K-means clustering analysis was useful in providing a clear and vivid picture of how the vowels were under each cluster, depending on their acoustic content, which further showed systematic deviation of the two varieties. These clusters which occurred throughout the analysis abide by the current phonological difference between high and low vowels as supported by the prior literature (Kashifa et al., 2025). The clustering also identified the different values of heights and backness vowels in ArE as opposed to PakE that displays the phonetic variance between the two variety. Lastly, the Logistic Regression classification has contributed additional evidence to the findings of the study, and the model could classify both ArE and PakE vowels, yet could not do the same with PakE vowels. It is an indicator of the fact that the acoustic divergence was greater in the ArE vowels, as ArE vowel tended to diverge on their targets more aptly to findings achieved by Alshangiti and Evans (2024), when the vowels voiced by the Arabic speakers of English were more centralised than those of native speakers of English.

This work can be applied in ESL pronunciation instruction, localisation to text-to-speech (TTS) and local intelligibility. Regional difference in the systematic variation in the production of vowels in PakE and ArE speakers can be better informed to inform certain teaching requirements. As an illustration, an instructor is allowed to highlight the distinctions between backness and frontness that characterize the two types. This study confirms the fact of systematic difference in the vowels between PakE and ArE and the two forms of the language are read abnormally in relation to RP especially in the central vowels and back vowels. Theoretical consequences of the current research can be used to authenticate the peculiarities of non-native forms of English, including PakE and ArE, their systematic quality in phonology, and the effect of L1 phonology on the production of vowels. But regarding limitations, findings are generalised because the sample is small, and the focus of the results is laid on the controlled speech. The limitations are surmountable in future studies as mentioned here by expanding the sample size and observing uncontrolled speech, which is likely to give more definite results on the vowel variations. Further studies can also be implemented in the future to use deep learning models to further explore vowel recognition and classification, in addition to other phonetic attributes, e.g., consonant production and intonation. This recommendation will come in handy in order to learn more about the regional version of English and enhance the accuracy of the phonetic analysis model.

Author Contributions: Nadia Safeer was the sole author of the manuscript, as it was conceptualised, designed, analysed and written. The author performed all the elements of the research, such as the methodology, software implementation, data curation, formal analysis, visualisation, drafting, reviewing, and editing independently. The author takes responsibility to the entire content of the manuscript and states that the work had been written entirely and did not involve any co-authorship. The author has read and accepted the published edition of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: These are available upon request.

Acknowledgments: During the preparation of this manuscript/study, the author used [GPT] for the purposes of [an outline]. The author has reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: “The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

References

1. Abbasi, A., Mansoor Channa, Memon, M., John, S., Ahmed, I., & Kumar, K. (2018). Acoustic characteristics of Pakistani English vowel sounds. *International Journal of English Linguistics*, 8(5), 27–27. <https://doi.org/10.5539/ijel.v8n5p27>
2. Abdelgadir, E. M. (2021). A contrastive study of English and Arabic vowel phonemes. *English Language Teaching*, 14(5), 58–65. <https://doi.org/10.5539/elt.v14n5p58>
3. Akewula, A. O. (2025). Exploring phonetic variations in Arabic and English: A study of minimal pairs on articulatory features. <https://doi.org/10.2139/ssrn.5230523>
4. Alshangiti, W., & Evans, B. G. (2024). Learning English vowels: The effects of different phonetic training modes on Arabic learners' production and perception. *The Journal of the Acoustical Society of America*, 156(1), 284–298. <https://doi.org/10.1121/10.0026451>
5. Bilal, H. A., Azher, M., Ishfaq, M., & Mumtaz, A. (2021). Acoustic investigation of back vowels of Pakistani English. *Review of Education, Administration & LAW*, 4(1), 37–52. <https://doi.org/10.47067/real.v4i1.110>
6. Colantoni, L., Kochetov, A., & Steele, J. (2023). L1 influence on the L2 acquisition of English word-final nasal place contrasts: An electropalatographic study of L1 Japanese and Spanish learners. *Laboratory Phonology*, 14(1). <https://doi.org/10.16995/labphon.6434>
7. Ghafoor, A., Afshan Gul Shahzadi, & Lubna Shahzadi. (2024). Exploring the phonological analysis of vowel sounds among speakers of Pakistani English. *Shnakhat*, 3(2), 311–322. <https://shnakhat.com/index.php/shnakhat/article/view/305>
8. Gusdian, R. I., & Lestiono, R. (2021). English and Arabic vowels: Ferreting out the similarity for bridging pronunciation accuracy. *Journal of English Language Teaching and Linguistics*, 6(2), 297–307. <https://doi.org/10.21462/jeltl.v6i2.544>
9. Hussain, S., Anjum, U., Safeer, N., & Malik, S. (2022). Acoustic analysis of English vowel sounds produced by Sindhi speakers. *Pakistan Journal of Society, Education and Language (PJSSEL)*, 9(1), 353–365. <https://jehanf.com/pjsel/index.php/journal/article/view/1027>
10. Kashifa, A., Safeer, N., Mubeen, A., & Sidrat-ul-Muntaha, S. (2025). Vowel duration and L1 influence in Pakistani English: An acoustic analysis of English monophthongs. *Journal of Applied Linguistics and TESOL (JALT)*, 8(1), 850–863. <https://jalt.com.pk/index.php/jalt/article/view/399>
11. Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588486>
12. Liu, Y., & Luo, Q. (2024). Deep machine learning-based analysis for intelligent phonetic language recognition. *Scalable Computing Practice and Experience*, 25(3), 1557–1563. <https://doi.org/10.12694/scpe.v25i3.2710>
13. Liu, Z. (2023). A systematic review: Exploring L2 vowel production from revised speech learning theory perspective. *Open Journal of Social Sciences*, 11(06), 452–463. <https://doi.org/10.4236/jss.2023.116029>

14. Malakar, M., & Keskar, R. B. (2021). Progress of machine learning based automatic phoneme recognition and its prospect. *Speech Communication*, 135, 37–53. <https://doi.org/10.1016/j.specom.2021.09.006>
15. Martinez, R. M., Goad, H., & Dow, M. (2021). L1 phonological effects on L2 (non-)naïve perception: A cross-language investigation of the oral–nasal vowel contrast in Brazilian Portuguese. *Second Language Research*, 39(2), 387–423. <https://doi.org/10.1177/026765832111044953>
16. Safeer, N., Anjum, U., & Saleem, T. (2024). Gender-based study of paired monophthongs: A sociophonetics approach. *3L: Language, Linguistics and Literature*, 30(2), 231–262. <https://doi.org/10.17576/3L-2024-3002-15>
17. Safeer, N., Malik, S., & Anjum, U. (2023). A descriptive analysis of English vowel sounds by L1 Pahari learners. *Pakistan Journal of Society, Education and Language (PJSEL)*, 9(2), 26–42. <https://jehanf.com/pjsel/index.php/journal/article/view/1119>
18. Turner, J. (2022). Phonetic development of an L2 vowel system and tandem drift in the L1: A residence abroad and L1 re-immersion study. *Language and Speech*, 66(3), 756–785. <https://doi.org/10.1177/00238309221133100>
19. Yaqub, M., & Khan, M. K. (2024). A descriptive study of vowels in dialects of Pakistani English. *The Journal of the Acoustical Society of America*, 155(3_Supplement), A169–A169. <https://doi.org/10.1121/10.0027193>

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.