Linguistic Forum

Published by
MARS Publishers

ISSN (Online) 2707-5273
Volume 7, Issue 3, 2025
http://doi.org/ 10.5281/zenodo. 21212705
Pages 01-22

Gender Differences in the Acoustic Properties of English Monophthongs Produced by Saraiki ESL Learners in Sindh

Research Article

Correspondence:

Abdul Manan
<mananjatoi8@gmail.com>
BS English Student, Department of English, Sindh Madressatul Islam University Karachi, Sindh, Pakistan.

Wafa Mansoor Buriro
<wafa.mansoor@smiu.edu.pk>
Assistant Professor, Department of English, Sindh Madressatul Islam University Karachi Sindh, Pakistan.

Muhammad Irfan Khan
<irfankhansir@gmail.com>
MS Linguistics Student - Department of English, Sindh Madressatul Islam University Karachi Sindh, Pakistan.

Publication Details

Received: September 25, 2025 **Accepted:** November 20, 2025 **Published:** December 6, 2025

Abstract

This study investigates gender-based differences in the acoustic properties of speech produced by Saraiki male and female English as a Second Language (ESL) learners. Specifically, it examines how vowel quality (F1 and F2), pitch (F0), and duration of English vowels differ between non-native Saraiki ESL learners and native English speakers. The eight English monophthongs selected for analysis were /iː/, /uː/, /ɑː/, /ɜː/, /ɔː/, /ɪ/, /ʌ/, and /ʊ/. Data were collected from 20 Sindhi-dialect Saraiki ESL learners, comprising 10 male and 10 female speakers. The recordings were analyzed using Praat software to measure formant frequencies (F1 and F2), pitch (F0), and vowel duration. The findings reveal significant gender-based differences, with female speakers exhibiting higher pitch (F0) and longer durations for short vowels than male speakers. The results also highlight the influence of the first language (L1) on second language (L2) vowel production, as Saraiki possesses articulatory patterns that differ from those of English. Overall, the findings

suggest that both physiological and sociolinguistic factors contribute to gender-based variation in English vowel production among Saraiki ESL learners, providing valuable insights into the role of gender in second language acquisition.

Keywords: acoustic phonetics; English monophthongs; gender differences; Saraiki ESL learners; vowel production

1. Introduction

In an increasingly interconnected world, understanding how L2 speakers produce and perceive English sounds is essential for improving language teaching methodologies and facilitating cross-cultural communication. The acoustic analysis of speech sounds enables researchers to gain deeper insights into the mechanisms underlying speech production. This area of research is particularly important for second-language learners, who often exhibit differences in the production of speech sounds because of the influence of their L1. The present study investigates gender-based differences in the production of English monophthongs by Saraiki speakers living in Sindh, whose first language is Saraiki but whose speech is influenced by the Sindhi dialect. In this linguistic context, Sindhi has influenced the Saraiki spoken by these learners, and this linguistic interference is expected to affect their production of English vowels. As noted by Derakhshan and Karimi (2015), second-language learners frequently transfer the linguistic forms, meanings, and cultural patterns of their first language while communicating in a second language.

Gender-based variation in speech production has attracted considerable attention because males and females may produce speech differently owing to physiological, psychological, and sociolinguistic factors. These differences become even more pronounced in second-language acquisition. According to Edwar et al. (2021), men tend to use more non-standard linguistic variants, whereas women are generally more likely to adopt prestige or standard variants. The authors attribute this tendency to sociolinguistic factors, including differences in access to educational, economic, and social resources. In addition to these social influences, physiological differences between males and females, particularly variations in vocal fold length and vocal tract anatomy, affect resonance, pitch, and stress patterns, thereby contributing to differences in vowel pronunciation.

Vowel production is primarily determined by the movement and position of the tongue within the oral cavity during articulation. In acoustic phonetics, vowel quality is commonly described using the first two formant frequencies, F1 and F2, which correspond to tongue height and tongue advancement, respectively.

The first formant frequency (F1) reflects the vertical position of the tongue during vowel production. High or close vowels are produced when the tongue is positioned near the roof of the mouth, creating a relatively narrow passage for airflow and resulting in lower F1 values. In contrast, low or open vowels are articulated with the tongue positioned lower in the oral cavity, creating a wider passage for airflow and producing higher F1 values. As explained by Jan Pýcha (2016), vowels articulated with a raised tongue position are classified as high (close) vowels, whereas those produced with a lowered tongue position are classified as low (open) vowels.

The second formant frequency (F2) represents the horizontal position of the tongue and is used to distinguish front and back vowels. Front vowels are produced when the front portion of the tongue is raised toward the alveolar ridge, resulting in higher F2 values. Conversely, back vowels are articulated with the tongue positioned toward the soft palate, producing lower F2 values. Pýcha (2016) further explained that vowels produced with the front portion of the tongue elevated are categorized as front vowels, whereas those articulated with the tongue positioned toward the back of the oral cavity are classified as back vowels.

Fundamental frequency (F0), commonly referred to as pitch, determines the perceived intonation or the rise and fall of speech. It reflects the rate at which the vocal folds vibrate during phonation. A greater number of vocal fold vibrations per second results in a higher pitch, whereas fewer vibrations produce a lower pitch. Because vocal fold anatomy differs between males and females, fundamental frequency is considered one of the most reliable acoustic indicators of gender-related variation in speech production.

1.1 Background of the Study

The Sindhi dialect of the Saraiki language is spoken in the Sindh province of Pakistan and in parts of India. Its vowel system shares certain similarities with that of English; however, several vowels differ in their articulatory and acoustic characteristics. For example, the Saraiki vowel /i/ closely corresponds to the English vowel /i/, whereas the Saraiki vowel /a/ is often realized similarly to the English vowel /æ/. These phonetic differences may create challenges for Sindhi-dialect Saraiki speakers when producing English monophthongs accurately. Previous research has shown that second-language (L2) learners frequently experience difficulty producing speech sounds that are absent from or differ substantially from those in their native language (L1) phonological system (Foley & Lahiri, 1991). This phenomenon is explained by the Perceptual Assimilation Model (Best, 1995), which proposes that L2 learners perceive and produce unfamiliar speech sounds by assimilating them into existing L1 phonetic categories, often resulting in non-native patterns of pronunciation. Furthermore, Sindhi speakers have been reported to produce certain vowels with relatively longer durations, reflecting language-specific phonological and phonetic characteristics (Kashifa et al., 2025).

Gender is another important factor influencing second-language speech production. Previous studies have suggested that males and females differ in the way they learn, acquire, and use language. Research on the relationship between gender and language use indicates that women often exhibit linguistic behaviors that differ from those of men (Mahmoo et al., 2015). Within the sociocultural context of the Indian subcontinent, women have traditionally had comparatively fewer opportunities to interact with speakers of different languages because of prevailing social norms and expectations. In contrast, men generally experience greater social mobility and broader exposure to multilingual environments. These differences in linguistic exposure may influence second-language proficiency and pronunciation patterns. Consequently, male and female Sindhi-dialect Saraiki speakers may differ in their production of English vowels. In addition to sociolinguistic factors, physiological differences—particularly variations in vocal fold length and vocal tract anatomy, also contribute to gender-based differences in vowel production, pitch, and other acoustic characteristics of speech.

1.2 Research Objectives

This study aims to achieve the following objectives:

1. To examine gender-based differences in the durational values of English monophthongs produced by male and female Sindhi-dialect Saraiki speakers.
2. To investigate differences in vowel quality (F1 and F2) and pitch (F0) of English monophthongs produced by male and female Sindhi-dialect Saraiki speakers.
3. To compare the vowel quality (F1 and F2) and duration of English monophthongs produced by native English speakers and Sindhi-dialect Saraiki speakers.

1.3 Research Questions

The study addresses the following research questions:

1. How do the durational values of English monophthongs differ between male and female Sindhi-dialect Saraiki speakers?
2. How do vowel quality (F1 and F2) and pitch (F0) of English monophthongs vary between male and female Sindhi-dialect Saraiki speakers?
3. How do vowel quality (F1 and F2) and duration of English monophthongs differ between native English speakers and Sindhi-dialect Saraiki speakers?

2. Literature Review

Acoustic phonetics is the branch of phonetics that examines the physical properties of speech sounds as they travel from the speaker's vocal tract to the listener's auditory system. The study of acoustic phonetics provides a framework for understanding how articulatory gestures are translated into acoustic signals and how these signals are perceived by listeners (Moore & Skidmore, 2019). This field enables researchers to employ advanced analytical techniques, including spectrograms and formant analysis, to investigate speech sounds with greater precision. A spectrogram is a visual representation of speech over time, displaying the distribution of acoustic energy across frequencies. According to V. Zue, the spectrogram is one of the most widely used tools for visualizing speech, while formant frequencies represent the resonant frequencies of the vocal tract that define vowel quality. Yaqub (2025) further reported that speech patterns are strongly influenced by gender. Male speakers generally exhibit a smaller vowel space (approximately 2.5 z-units²) than female speakers (approximately 4.5 z-units²). The larger vowel space observed among females indicates greater articulatory variation and clearer vowel distinctions, whereas male speech tends to be comparatively compressed.

The analysis of formant frequencies (F1 and F2) constitutes the foundation of acoustic studies of vowel production. Aalto (2018) explained that formants are the dominant frequency components of speech that appear as local maxima in the power spectrum of an acoustic signal. These resonant frequencies arise from the resonance characteristics of the vocal tract and are particularly important in distinguishing vowel sounds. Formant frequencies provide valuable information about tongue position during articulation. According to Roach, the first formant (F1) primarily reflects vowel

height, whereas the second formant (F2) represents the frontness or backness of the tongue. Lower F1 values correspond to high (close) vowels because the tongue is positioned near the roof of the mouth, reducing the oral cavity and restricting airflow. Conversely, higher F1 values are associated with low (open) vowels, where the tongue occupies a lower position and creates a larger oral cavity. Similarly, higher F2 values indicate front vowels, while lower F2 values correspond to back vowels. Safeer et al. (2024) also reported significant gender-based differences in the acoustic characteristics of Pakistani English monophthongs. Their findings demonstrated that female speakers produced significantly longer vowel durations than male speakers, highlighting the influence of gender on vowel quality and temporal characteristics.

These acoustic properties are fundamental not only to understanding vowel production but also to explaining vowel perception. The relationship between speech production and perception becomes particularly important in second-language (L2) acquisition. Ali (2013) reported that English language learners experience considerable difficulty in acquiring English vowels because of interference from their first language (L1). Learners often substitute unfamiliar English vowels with acoustically or articulatorily similar vowels from their native language. Similarly, Abbasi et al. (2025) found that Pakistani English speakers transferred Voice Onset Time (VOT) patterns from their first language to English voiced stops. Their study further demonstrated that both gender and place of articulation significantly influenced these acoustic features. The stimuli included Sindhi L1 aspirated plosives (/pʰ/, /ʈʰ/, and /kʰ/), voiced plosives (/b/, /d/, and /ɡ/), and their English L2 counterparts.

Saraiki is spoken by approximately 15 million people across Pakistan and parts of India and belongs to the Indo-Aryan language family (Khan & Rashid, 2010). The language possesses a rich vowel inventory characterized by both oral and nasal vowels. Malghani (2019) reported that Saraiki contains 17 vowels, including nine oral vowels and eight nasal vowels. In addition, Saraiki consists of several regional dialects, including Southern Saraiki, Northern Saraiki, Jhangi Saraiki, Shahpuri Saraiki, Central Saraiki, and Sindhi Saraiki (Shackle, as cited in Malghani, 2019). Although Sindhi-dialect Saraiki shares several phonological features with other Saraiki varieties, it also demonstrates important phonemic differences resulting from prolonged contact with the Sindhi language. For example, Sindhi contains approximately 43 phonemes, whereas Saraiki contains approximately 48 phonemes (Malghani, 2019). These structural differences create a complex phonological environment for Saraiki learners when acquiring English vowel contrasts.

The English vowel system differs substantially from that of Saraiki. Standard English consists of approximately 44 phonemes, including 24 consonants and 20 vowels, two of which function as semivowels. The vowel inventory comprises 12 monophthongs and eight diphthongs. Unlike Saraiki, English lacks phonemic nasal vowels, creating additional pronunciation challenges for Saraiki learners. Alam and Khan (2015) observed that Saraiki learners frequently substitute English monophthongs with acoustically similar diphthongs or gliding vowels. For example, the English vowel /æ/ is often realized as /iː/. Similarly, Joyo and Memon (2023) investigated phonological variation among male and female Sindhi learners of English. Their findings revealed relatively minor differences in the production of vowels such as /ʌ/, /ə/, /e/, /æ/, /ɜː/, and /ɑː/, whereas greater variation was observed in vowels and diphthongs including /ɒ/, /ʊ/, /eɪ/, /əɪ/, /ɪə/, /iː/, /eə/, /ɔː/, /aʊ/, /aɪ/, /əʊ/, /uː/, and /ʊə/.

Gender further complicates the production of English vowels because males and females differ anatomically in ways that influence speech acoustics. Female speakers generally possess shorter vocal tracts and shorter vocal folds than male speakers. Markova (2016) explained that longer vocal folds vibrate at lower frequencies, producing lower fundamental frequency (F0), whereas shorter vocal folds vibrate more rapidly and generate higher pitch values. Goldstein (1980) reported that the average vocal fold length is approximately 14.1 mm for females and 17 mm for males. Zhang (2021) further argued that differences in vocal fold size influence their vibratory behavior, thereby contributing to gender-based differences in speech acoustics.

In addition to biological differences, sociolinguistic factors also influence gender-based variation in second-language pronunciation. In many Sindhi-speaking regions where the Sindhi dialect of Saraiki is spoken, males generally have greater access to English-medium education than females (Muneer, 2021). Cultural practices often prioritize boys' education over girls', resulting in greater multilingual exposure for males. Differences in educational opportunities and language exposure may consequently influence the acquisition and pronunciation of English speech sounds. However, the relationship between social exposure and pronunciation proficiency remains complex and requires empirical investigation rather than broad generalization.

Despite extensive research on second-language acquisition and the acoustic characteristics of speech, relatively little attention has been paid to gender-based variation in English vowel production among Sindhi-dialect Saraiki ESL learners. This gap is particularly evident in studies examining the combined effects of gender, first-language interference, vowel quality, pitch, and duration. The present study addresses this gap by investigating the acoustic characteristics of English monophthongs produced by male and female Sindhi-dialect Saraiki learners.

3. Methodology

3.1 Sampling

The participants for this study were undergraduate and postgraduate students enrolled at Mehran University of Engineering and Technology, University of Sindh, Quaid-i-Azam University, and Quaid-e-Awam University of Engineering, Science and Technology. The participants represented various academic disciplines, including BS English, BS Urdu, BS Chemistry, BS Civil Engineering, and MS Mathematics. All participants were at least 18 years of age.

The study employed purposive sampling to select 20 Sindhi-dialect Saraiki speakers, comprising 10 male and 10 female English as ESL learners. This balanced sample enabled a gender-based comparison of vowel quality, vowel duration, and pitch in the production of selected English monophthongs.

3.2 Data Collection

The speech corpus consisted of eight English words representing eight English monophthongs: /iː/, /uː/, /ɑː/, /ɜː/, /ɔː/, /ɪ/, /ʌ/, and /ʊ/. All target vowels occurred in consonant–vowel–consonant (CVC) word structures. The selected vowels included both long and short

monophthongs. According to Roach (2024), the English long vowels are /iː/, /uː/, /ɑː/, /ɜː/, and /ɔː/, which are represented by a vowel symbol followed by the length mark (/ː/). The short vowels included in the study were /ɪ/, /ʌ/, and /ʊ/.

One lexical item was selected to represent each target vowel: treat (/iː/), rude (/uː/), bar (/ɑː/), bird (/ɜː/), cause (/ɔː/), fit (/ɪ/), truck (/ʌ/), and wood (/ʊ/). The word list was distributed to all participants as a PDF document via WhatsApp. Participants were instructed to record each word individually in isolation to minimize coarticulatory effects and to ensure optimal recording quality. Each word was pronounced three times by every participant.

Table 1: Speech Corpus

Vowel	Word
/iː/	Treat
/uː/	Rude
/ɑː/	Bar
/ɜː/	Bird
/ɔː/	Cause
/ɪ/	Fit
/ʌ/	Truck
/ʊ/	Wood

Note. Each lexical item represents one target English monophthong analyzed in the study.

3.3 Speech Recording and Acoustic Analysis

Speech recordings were obtained from the 20 Sindhi-dialect Saraiki speakers using the voice recorder of a mobile device. Participants were instructed to complete the recordings in a quiet environment to minimize background noise and ensure high-quality audio.

Each participant produced eight target words three times, resulting in a total of 480 speech samples (8 words × 20 participants × 3 repetitions). The recordings were converted into MP3 format and imported into Praat for acoustic analysis.

Within Praat, the vowel segment of each recording was manually identified and isolated. The View & Edit function was used to display the waveform and spectrogram for each speech sample. Acoustic measurements included the first formant (F1), second formant (F2), fundamental frequency (F0), and vowel duration. Formant frequencies (F1 and F2) and fundamental frequency (F0) were measured in hertz (Hz), whereas vowel duration was initially measured in seconds and subsequently converted to milliseconds (ms) for statistical analysis.

4. Results

This section presents the acoustic analysis of the selected English monophthongs produced by Sindhi-dialect Saraiki English as a Second Language (ESL) learners. The analysis focuses on the first formant (F1), second formant (F2), fundamental frequency (F0), and vowel duration. As each participant produced every target

word three times, the reported values represent the arithmetic mean of the three repetitions. Averaging the acoustic measurements minimizes variability across repetitions and provides a more reliable representation of participants' vowel production.

The reference values for native English speakers, including average F1, F2, F0, and vowel duration, were obtained from the study *Acoustic Characteristics of American English Vowels*.

4.1 Vowel /iː/

The vowel /iː/ is a long front unrounded monophthong produced with the tongue positioned high and toward the front of the oral cavity, while the lips remain slightly spread (Roach, 2024). Table 2 presents the average values of F0, F1, F2, and vowel duration for native and non-native speakers across gender.

Table 2: Acoustic Characteristics of the Vowel /iː/

Gender	F0 (Hz)	F1 (Hz)	F2 (Hz)	Duration (ms)
Non-native female	454	242	2443	190
Native female	227	437	2761	306
Non-native male	128	380	2292	175
Native male	138	342	2322	243

4.2 Vowel /uː/

The vowel /uː/ is a long back rounded monophthong articulated with the tongue positioned high and toward the back of the mouth near the soft palate. It is classified as a tense back rounded vowel (Abbasi et al., 2022). Table 3 presents the average acoustic measurements for native and non-native speakers.

Table 3: Acoustic Characteristics of the Vowel /uː/

Gender	F0 (Hz)	F1 (Hz)	F2 (Hz)	Duration (ms)
Non-native female	235	377	1067	198
Native female	235	459	1105	303
Non-native male	164	467	1013	222
Native male	143	378	997	237

4.3 Vowel /ɑː/

The vowel /ɑː/ is a long back unrounded monophthong articulated with the tongue positioned low and toward the back of the oral cavity. During its production, the lips remain unrounded while the dorsum of the tongue moves slightly backward (Demirezen, 2020). Table 4 presents the acoustic measurements.

Table 4**:** Acoustic Characteristics of the Vowel /ɑː/

Gender	F0 (Hz)	F1 (Hz)	F2 (Hz)	Duration (ms)
Non-native female	204	793	1368	255
Native female	215	936	1551	323
Non-native male	199	616	1258	260
Native male	123	768	1333	267

4.4 Vowel /ɜː/

The vowel /ɜː/ is a long mid-central unrounded vowel produced with the tongue positioned centrally at an intermediate height. During articulation, the tongue rises toward the palate without reaching either the front or back extremes (Roach, 2001). Table 5 summarizes the acoustic measurements.

Table 5**:** Acoustic Characteristics of the Vowel /ɜː/

Gender	F0 (Hz)	F1 (Hz)	F2 (Hz)	Duration (ms)
Non-native female	294	603	1570	249
Native female	217	523	1588	321
Non-native male	122	560	1575	241
Native male	130	474	1379	263

4.5 Vowel /ɔː/

The vowel /ɔː/ is a long mid-back rounded monophthong produced with the back of the tongue raised to a mid position while the lips remain rounded during articulation (Pratiwi & Indrayani, 2021). Table 6 presents the average acoustic measurements.

Table 6: Acoustic Characteristics of the Vowel /ɔː/

Gender	F0 (Hz)	F1 (Hz)	F2 (Hz)	Duration (ms)
Non-native female	223	552	1276	226
Native female	210	781	1136	353
Non-native male	116	596	1186	275
Native male	121	652	997	283

4.6 Vowel /ɪ/

The vowel /ɪ/ is a short near-close front unrounded vowel produced with the tongue raised toward the front of the oral cavity while the lips remain relaxed and unrounded. Table 7 presents the average acoustic measurements.

Table 7**:** Acoustic Characteristics of the Vowel /ɪ/

Gender	F0 (Hz)	F1 (Hz)	F2 (Hz)	Duration (ms)
Non-native female	245	458	2029	369
Native female	224	483	2365	237
Non-native male	130	463	2106	88
Native male	135	427	2034	192

4.7 Vowel /ʌ/

The vowel /ʌ/ is a short open-mid back unrounded vowel articulated with the tongue positioned centrally toward the back of the oral cavity while the lips remain neutral. Table 8 presents the acoustic measurements.

Table 8: Acoustic Characteristics of the Vowel /ʌ/

Gender	F0 (Hz)	F1 (Hz)	F2 (Hz)	Duration (ms)
Non-native female	225	669	1503	405
Native female	218	753	1426	226
Non-native male	125	611	1323	103
Native male	133	623	1200	188

4.8 Vowel /ʊ/

The vowel /ʊ/ is a short near-close back rounded monophthong produced with the tongue raised toward the back of the oral cavity while the lips remain rounded throughout articulation. Table 9 presents the average acoustic measurements.

Table 9: Acoustic Characteristics of the Vowel /ʊ/

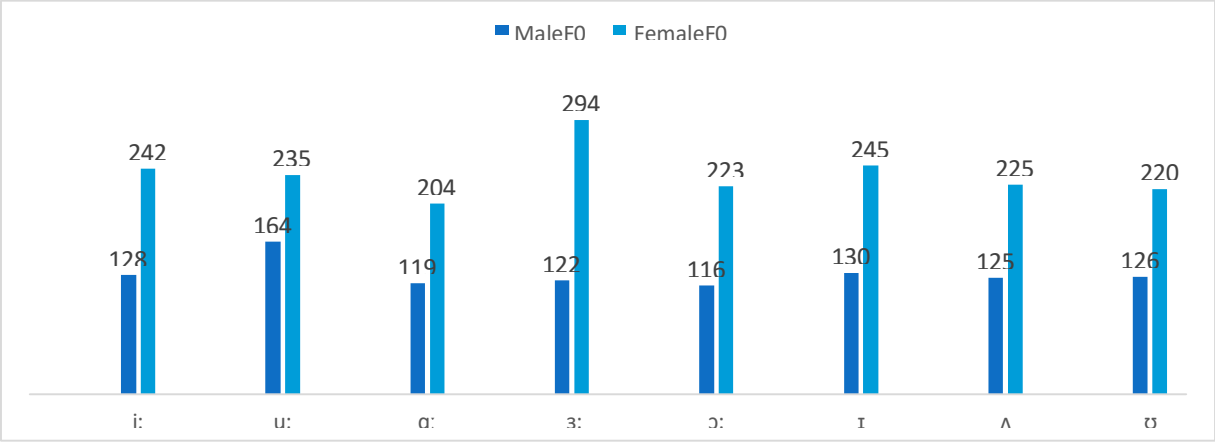
Gender	F0 (Hz)	F1 (Hz)	F2 (Hz)	Duration (ms)
Non-native female	220	433	1177	367
Native female	230	519	1225	249
Non-native male	126	505	1139	104
Native male	133	469	1122	192

The acoustic analysis demonstrates measurable differences in the fundamental frequency (F0), vowel quality (F1 and F2), and vowel duration between male and female Sindhi-dialect Saraiki ESL learners. Comparisons with native English speakers further indicate variations in vowel production that may be attributed to gender-related physiological characteristics and first-language influence. These findings are discussed in greater detail in the subsequent discussion section.

5. Discussion

The gender differences in the acoustic properties of English monophthongs produced by Sindhi-dialect Saraiki ESL learners are discussed in this section. The differences in formant frequencies (F1 and F2), vowel duration, and pitch (F0) were analyzed to identify variations between male and female speakers. The findings are discussed in terms of the average values of F1, F2, F0, and vowel duration. The average values of the fundamental frequency (F0) for male and female speakers are illustrated in Figure 1**.**

Figure 1: Average F0 Values of Male and Female Speakers



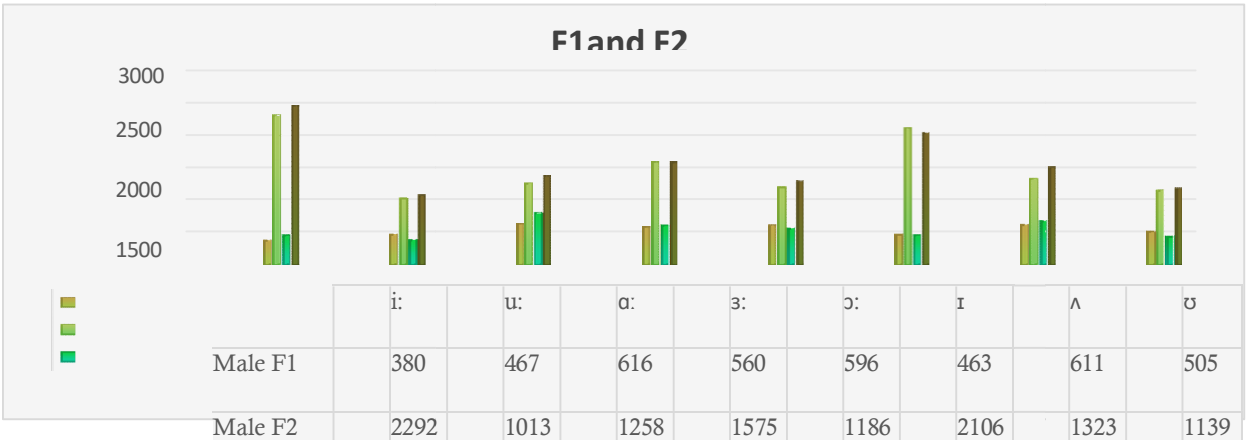
Fundamental frequency (F0) of sounds refers to the lowest frequency present in a complex sound wave. In speech, it is often referred to as pitch (Abbasi et al., 2021) and is determined by the vibration rate of the vocal cords (Zsiga, 2013). Fundamental frequency (F0) refers to the number of sound waves produced by the vocal cords in one second. The higher the F0 value, the faster the vocal cords vibrate, resulting in a greater number of sound waves produced per second. The analyzed data, illustrated in Figure 1, demonstrate that, for all the selected English monophthongs, the female Sindhi-dialect Saraiki speakers have higher F0 values than the male speakers. The average F0 values for the long vowels (/iː/, /uː/, /ɑː/, /ɜː/, and /ɔː/) produced by the female speakers range from 204 Hz to 294 Hz, whereas those of the male speakers range from 116 Hz to 164 Hz. Similarly, for the short vowels (/ɪ/, /ʌ/, and /ʊ/), the average F0 values of the male speakers range from 125 Hz to 130 Hz, while those of the female speakers range from 220 Hz to 245 Hz.

The pitch (F0) of the female speakers for the vowel /iː/ was measured at 242 Hz, whereas the male speakers produced an F0 value of 128 Hz for the same vowel. Similarly, the female speakers produced an F0 value of 235 Hz for the vowel /uː/, compared with 164 Hz for the male speakers. In addition, for the vowel /ɑː/, the female speakers produced an F0 value of 204 Hz, whereas the male speakers exhibited an F0 value of 199 Hz, indicating a slightly higher pitch for the female speakers. The difference between male and female speakers for the vowel /ɜː/ was more pronounced. The male speakers produced an F0 value of 122 Hz, whereas the female speakers produced an average F0 value of 294 Hz. Likewise, the female speakers produced a higher F0 value for the vowel /ɔː/ (223 Hz) than the male speakers (116 Hz).

Moreover, this pattern of higher F0 values among female speakers was also observed for the short vowels. For the vowel /ɪ/, the female speakers produced an average F0 value of 245 Hz, whereas the male speakers produced 130 Hz. Similarly, the female speakers produced an average F0 value of 225 Hz for the vowel /ʌ/, while the male speakers produced 125 Hz. The same pattern was observed for the vowel /ʊ/, where the male speakers produced an F0 value of 126 Hz, whereas the female speakers produced an F0 value of 220 Hz.

The analysis indicates that the F0 values are consistently higher for all vowels produced by the female speakers. These differences can be explained by physiological factors. The length of the vocal cords differs between males and females, which contributes to the observed variation in fundamental frequency. In adult speakers, the average vocal tract length measures approximately 16.9 cm in males and 14.1 cm in females (Goldstein, 1980). This anatomical difference generally results in lower formant and fundamental frequency values for male speakers than for female speakers. As Goldstein (1980) explained, the shorter vocal cords of female speakers vibrate more rapidly when air passes through them, resulting in higher F0 values. The average F1 and F2 values of the male and female speakers are illustrated in Figure 2.

Figure 2: Average F1 and F2 Values of Male and Female Speakers



	iː	uː	ɑː	ɜː	ɔː	ɪ	ʌ	ʊ
Male F1	380	467	616	560	596	463	611	505
Male F2	2292	1013	1258	1575	1186	2106	1323	1139
FemaleF1	454	377	793	603	552	458	669	433
FemaleF2	2443	1067	1368	1570	1276	2029	1503	1177

The data demonstrate that there are significant variations in the vowel quality of male and female speakers in the production of English monophthongs. The formant frequencies, F1 and F2, provide information about the height and advancement of the tongue, which differentiate various vowel sounds.

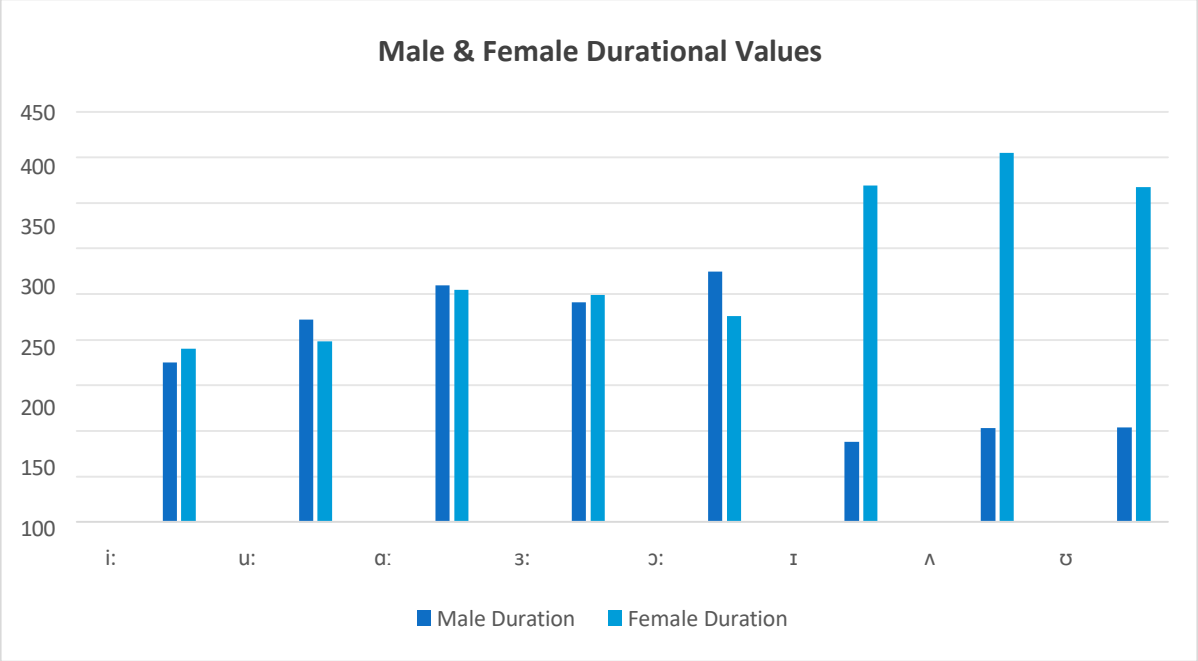
Among the Sindhi-dialect Saraiki ESL learners, there are significant differences in the acoustic properties of English monophthongs produced by male and female speakers. For the vowel /iː/, the female speakers had an average F1 value of 454 Hz and an F2 value of 2443 Hz. In contrast, the male speakers had an average F1 value of 380 Hz and an F2 value of 2292 Hz. Similarly, for the vowel /uː/, the female speakers had an average F1 value of 377 Hz and an F2 value of 1067

Hz, whereas the male speakers had an average F1 value of 467 Hz and an F2 value of 1013 Hz. For the vowel /ɑ:/, the female speakers had an average F1 value of 793 Hz and an F2 value of 1368 Hz, while the male speakers produced an average F1 value of 616 Hz and an F2 value of 1258 Hz.

This pattern also extended to the vowel /ɜ:/, where the female speakers had an average F1 value of 603 Hz and an F2 value of 1570 Hz, compared with the male speakers, who had an average F1 value of 560 Hz and an F2 value of 1575 Hz. Conversely, for the vowel /ɔ:/, the male speakers had a slightly higher F1 value (596 Hz) than the female speakers (552 Hz), whereas the female speakers had a higher F2 value (1276 Hz) than the male speakers (1186 Hz).

Similarly, for the vowel /ɪ/, the female speakers had an average F1 value of 458 Hz and an F2 value of 2029 Hz, whereas the male speakers recorded an average F1 value of 463 Hz and an F2 value of 2106 Hz. Thus, the male speakers exhibited slightly higher F1 and F2 values than the female Sindhi-dialect Saraiki speakers for this vowel. For the vowel /ʌ/, the female speakers had an average F1 value of 669 Hz and an F2 value of 1503 Hz. In contrast, the male speakers had an average F1 value of 611 Hz and an F2 value of 1323 Hz. For the vowel /ʊ/, the male speakers produced an average F1 value of 505 Hz and an F2 value of 1139 Hz, whereas the female speakers had an average F1 value of 433 Hz and an F2 value of 1177 Hz. The average durational values (in milliseconds) of the male and female speakers are illustrated in Figure 3.

Figure 3: Mean Durational Values of Vowels Produced by Male and Female Speakers



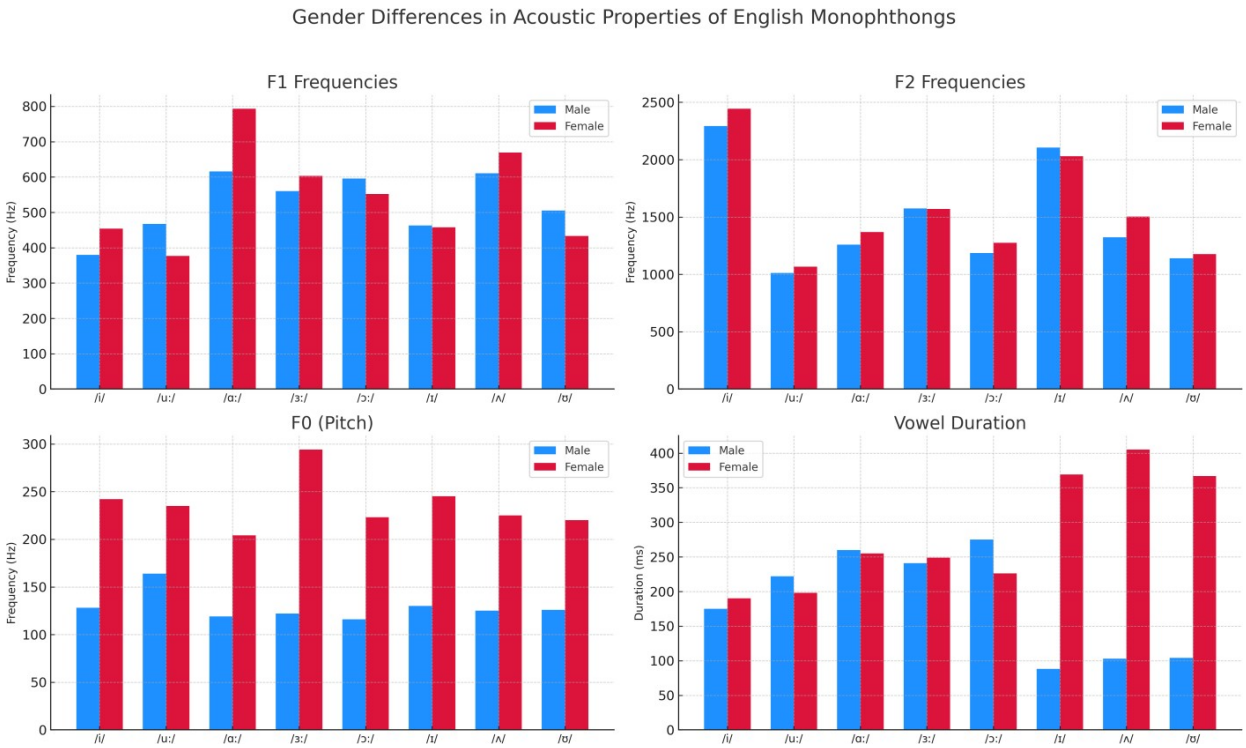
The data demonstrate that there are significant variations in the vowel quality of male and female speakers in the production of English monophthongs. The formant frequencies, F1 and F2, provide information about the height and advancement of the tongue, which differentiate various vowel sounds.

Among the Sindhi-dialect Saraiki ESL learners, there are significant differences in the acoustic properties of English monophthongs produced by male and female speakers. For the vowel /iː/, the female speakers had an average F1 value of 454 Hz and an F2 value of 2443 Hz. In contrast, the male speakers had an average F1 value of 380 Hz and an F2 value of 2292 Hz. Similarly, for the vowel /uː/, the female speakers had an average F1 value of 377 Hz and an F2 value of 1067 Hz, whereas the male speakers had an average F1 value of 467 Hz and an F2 value of 1013 Hz. For the vowel /ɑː/, the female speakers had an average F1 value of 793 Hz and an F2 value of 1368 Hz, while the male speakers produced an average F1 value of 616 Hz and an F2 value of 1258 Hz.

This pattern also extended to the vowel /ɜː/, where the female speakers had an average F1 value of 603 Hz and an F2 value of 1570 Hz, compared with the male speakers, who had an average F1 value of 560 Hz and an F2 value of 1575 Hz. Conversely, for the vowel /ɔː/, the male speakers had a slightly higher F1 value (596 Hz) than the female speakers (552 Hz), whereas the female speakers had a higher F2 value (1276 Hz) than the male speakers (1186 Hz).

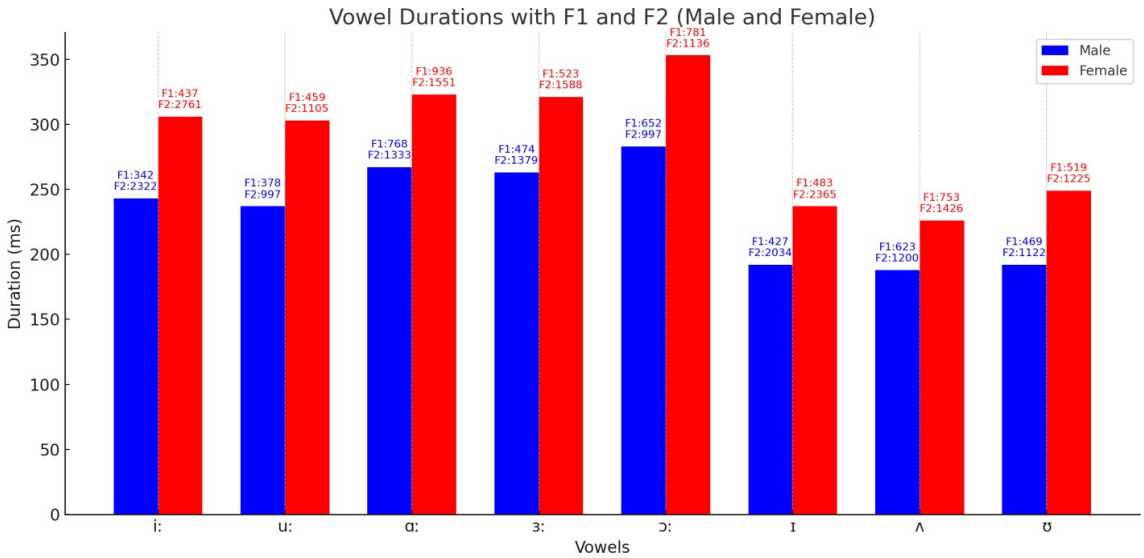
Similarly, for the vowel /ɪ/, the female speakers had an average F1 value of 458 Hz and an F2 value of 2029 Hz, whereas the male speakers recorded an average F1 value of 463 Hz and an F2 value of 2106 Hz. Thus, the male speakers exhibited slightly higher F1 and F2 values than the female Sindhi-dialect Saraiki speakers for this vowel. For the vowel /ʌ/, the female speakers had an average F1 value of 669 Hz and an F2 value of 1503 Hz. In contrast, the male speakers had an average F1 value of 611 Hz and an F2 value of 1323 Hz. For the vowel /ʊ/, the male speakers produced an average F1 value of 505 Hz and an F2 value of 1139 Hz, whereas the female speakers had an average F1 value of 433 Hz and an F2 value of 1177 Hz. The average durational values (in milliseconds) of the male and female speakers are illustrated in Figure 3.

Figure 3: Mean Durational Values of Vowels Produced by Male and Female Speakers



The vowel quality (F1 and F2) and duration of the English monophthongs were compared between native English speakers and Sindhi-dialect Saraiki speakers to examine differences in tongue height, tongue advancement, and vowel duration. Figure 5 illustrates the average formant frequencies (F1 and F2) and the average durational values (in milliseconds) of native English speakers.

Figure 5: Average F1, F2, and Durational Values of Native English Speakers



5.1 Formant frequencies (F1 and F2) between Saraiki ESL learners and native speakers

The formant frequencies (F1 and F2) of English monophthongs vary noticeably between non-native Sindhi-dialect Saraiki ESL learners and native English speakers. These differences reflect variations in the articulation of English vowels. The first formant frequency (F1) is associated with tongue height. When the tongue is positioned lower in the oral cavity, the F1 value is higher, resulting in the production of open vowels. Conversely, when the tongue is positioned higher, the F1 value is lower, resulting in the production of close vowels. The second formant frequency (F2) is associated with the horizontal position of the tongue and indicates the frontness or backness of a vowel. Higher F2 values are associated with front vowels, whereas lower F2 values indicate back vowels. Thus, a lower F1 value generally indicates that the tongue is positioned higher, producing a close vowel, whereas a higher F1 value suggests an open vowel. Similarly, a decrease in F2 corresponds to a vowel articulated farther back in the oral cavity (Roach, 2024, p. 42). Figure 6 illustrates the comparison of the F1 and F2 values between non-native male (Sindhi-dialect Saraiki) speakers and native male speakers.

Figure 6: Comparison of F1 and F2 Values Between Non-Native Male (Sindhi-Dialect Saraiki) and Native Male Speakers

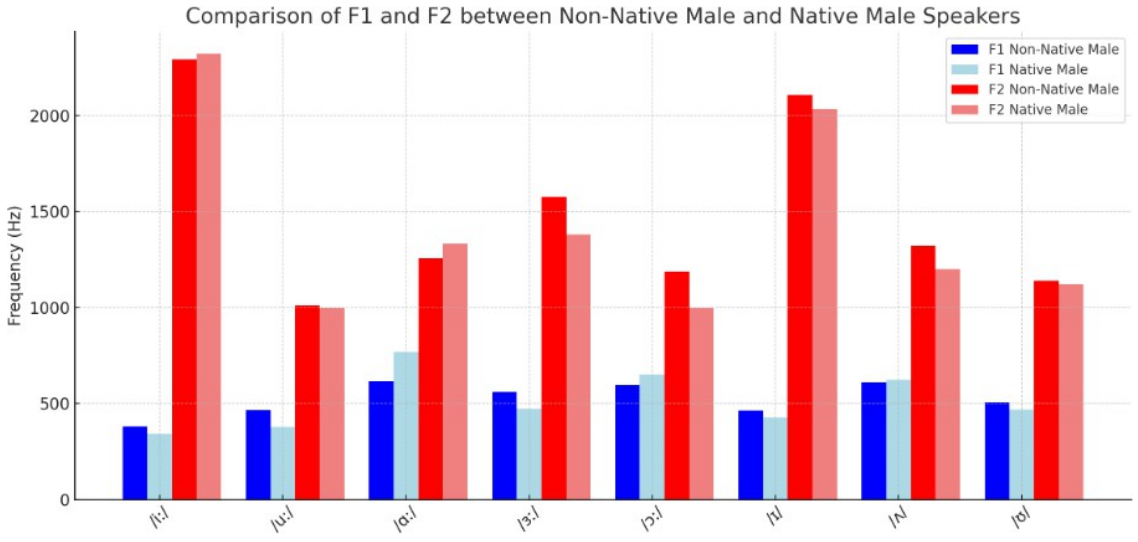
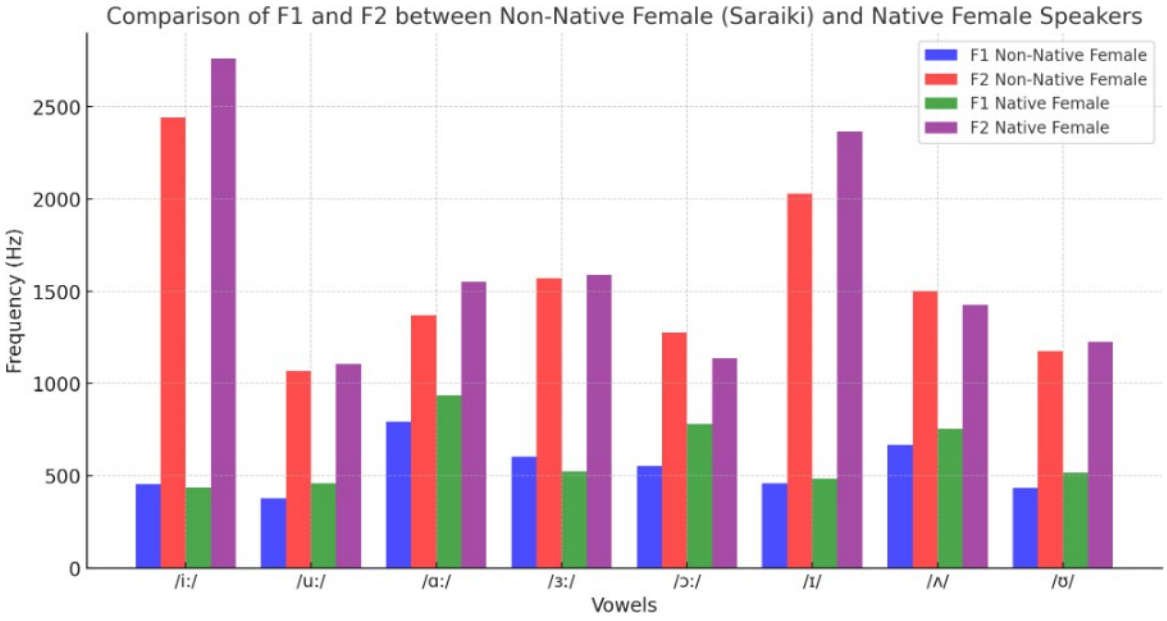


Figure 7 illustrates the comparison of the F1 and F2 values between non-native female (Sindhi-dialect Saraiki) speakers and native female speakers.

Figure 7: Comparison of F1 and F2 Values Between Non-Native Female (Sindhi-Dialect Saraiki) and Native Female Speakers



The Saraiki female speakers produced the vowel /iː/ with a more fronted articulation and higher tongue position than the native female speakers. Similarly, the Saraiki male speakers also articulated this vowel with a closer and more fronted tongue position than the native male speakers.

The formant frequencies (F1 and F2) of the non-native male and female speakers were compared with those of the native speakers for the vowel /uː/. The findings reveal that the Saraiki female speakers produced the vowel with the tongue positioned closer to the roof of the oral cavity and toward the back of the mouth. Consequently, both the F1 and F2 values were lower than those of the native female speakers. Conversely, the non-native male speakers produced the same vowel with higher F1 and F2 values than the native male speakers, indicating a more open tongue position with greater tongue advancement during vowel production.

For the vowel /ɑː/, the Saraiki male speakers produced the vowel with a more open and fronted tongue position in the oral cavity than the native male speakers. In contrast, the Saraiki female speakers produced the vowel with lower F1 and F2 values, indicating a higher and more retracted tongue position than the native female speakers. Similarly, for the vowel /ɜː/, the Saraiki female speakers exhibited higher F1 and F2 values, whereas the Saraiki male speakers had a higher F1 value and a lower F2 value than the native male speakers. These findings suggest that the non-native female speakers produced the vowel with a lower and more fronted tongue position, whereas the non-native male speakers articulated the same vowel with a lower but more retracted tongue position.

For the vowel /ɔː/, the Saraiki female speakers produced lower F1 and higher F2 values than the native female speakers, indicating a closer and more fronted tongue position during articulation. On the other hand, the Saraiki male speakers exhibited a more open and back tongue position in the oral cavity than the native male speakers.

In the case of the short vowels, the female Sindhi-dialect Saraiki ESL learners articulated the vowels /ɪ/ and /ʊ/ with the tongue raised closer to the alveolar ridge, resulting in lower F1 values. The lower F2 values for both vowels suggest that the tongue was positioned farther back in the oral cavity than that of the native female speakers.

Conversely, for the same vowels (/ɪ/ and /ʊ/), the Saraiki male speakers produced higher F1 and F2 values than the native male speakers, indicating relatively more open and fronted vowel articulation. Furthermore, both the non-native male and female speakers produced the vowel /ʌ/ with lower F1 and higher F2 values, indicating a relatively higher and more fronted tongue position. In contrast, the native male and female speakers exhibited higher F1 and lower F2 values, suggesting a lower and more retracted tongue position during the production of the same vowel.

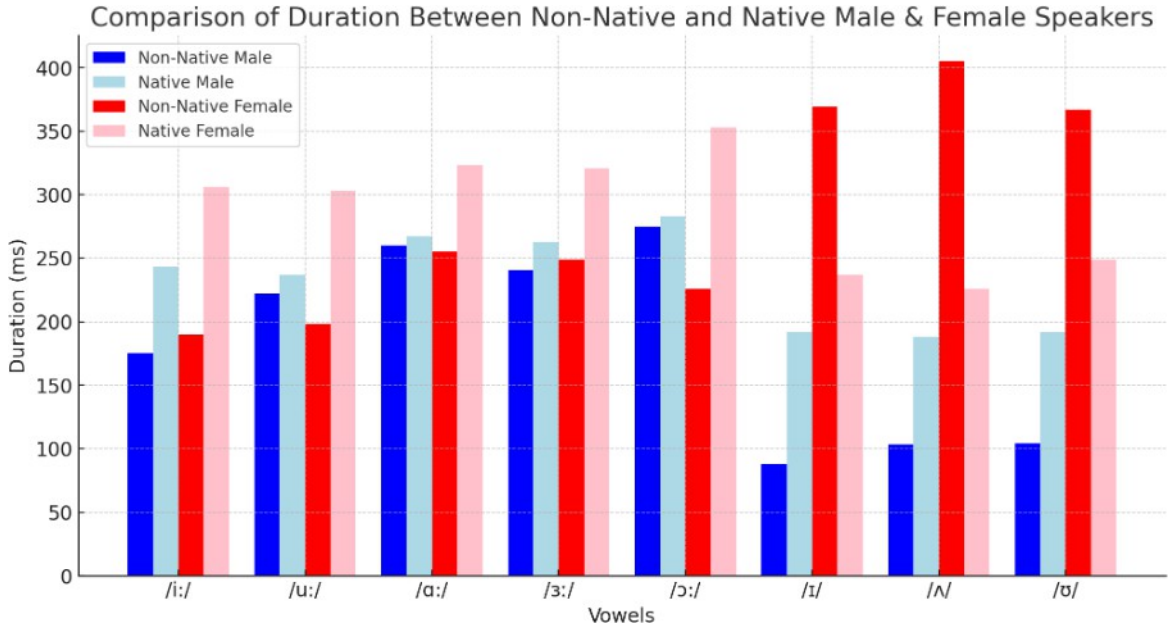
The findings further indicate that the male Saraiki ESL learners produced the vowels /ʊ/ and /uː/ with articulatory patterns that were relatively close to those of the native male speakers in terms of F2, with differences of 17 Hz and 16 Hz, respectively. Similarly, the Saraiki female speakers articulated the vowel /ɜː/ with an articulatory pattern that closely resembled that of the native female speakers, differing by only 18 Hz in F2. Furthermore, the F1 values of the Saraiki male speakers for the vowels /ɪ/, /ʊ/, and /ɜː/ differed from those of the native male speakers by only 5

Hz, 7 Hz, and 8 Hz, respectively, indicating only slight differences in pronunciation. Likewise, the Saraiki female speakers produced the vowel /iː/ with an F1 value differing by only 15 Hz from that of the native female speakers, suggesting a similar articulatory pattern.

5.2 Mean Durational Values between Saraiki ESL Learners and Native Speakers

Figure 8 illustrates the comparison of the mean durational values between native and non-native male and female speakers.

Figure 8: Mean Durational Values of Native and Non-Native Male and Female Speakers



A key difference is observed in the durational values of the non-native male and female speakers when compared with the native speakers. The study reveals that all of the long vowels (/iː/, /uː/, /ɑː/, /ɜː/, and /ɔː/) are pronounced with a shorter duration by the Saraiki female speakers than by the native female speakers. The most notable reductions in vowel duration are observed for the vowels /ɔː/ and /iː/, with differences of 127 ms and 116 ms, respectively.

Conversely, the short vowels (/ɪ/, /ʌ/, and /ʊ/) are pronounced with longer durations by the Saraiki female speakers than by the native female speakers. The greatest difference is observed for the vowel /ʌ/, where the Saraiki female speakers produced the vowel with a duration that was 179 ms longer than that of the native female speakers. Similarly, for the vowels /ɪ/ and /ʊ/, the Saraiki female speakers produced durations that were 132 ms and 118 ms longer, respectively, than those of the native female speakers.

The differences in the durational values of the non-native female speakers suggest that the Saraiki female speakers experience difficulty in maintaining the vowel length required for English monophthongs.

In contrast to the Saraiki female speakers, the non-native male speakers shortened almost all of the vowels, including both long and short vowels, when compared with the native male speakers. The greatest reductions were observed for the short vowels /ɪ/, /ʌ/, and /ʊ/. For example, the vowel /ɪ/ was shortened by 110 ms compared with the native male speakers. Similar reductions were observed for the vowels /ʌ/ and /ʊ/, with differences of 85 ms and 88 ms, respectively. Furthermore, the long vowels were also produced with shorter durations by the Saraiki male speakers than by the native male speakers. However, only a slight difference was observed for the vowel /ɔː/, with a difference of 8 ms. The results suggest that the Saraiki male speakers do not produce English monophthon

6. Conclusion

The study highlights notable gender-based differences in the production of English monophthongs among Sindhi-dialect Saraiki ESL learners. The female speakers exhibited higher pitch (F0) values for all of the selected vowels than the male speakers. These differences can be attributed to physiological factors, particularly the shorter vocal folds of female speakers, which vibrate more rapidly during speech production and consequently produce higher fundamental frequency (F0) values. In terms of vowel quality (F1 and F2), the female speakers generally produced higher formant frequencies, indicating relatively more fronted and open articulation than their male counterparts.

Moreover, significant differences were observed in the durational values of English monophthongs between the male and female speakers. The female speakers produced the short vowels (/ɪ/, /ʌ/, and /ʊ/) with longer durations than the male speakers. In contrast, the male speakers produced some of the long vowels (/uː/, /ɑː/, and /ɔː/) with slightly longer durations than the female speakers. However, both groups demonstrated noticeable differences from the native English speakers.

Furthermore, the comparison between native and non-native speakers revealed that the Sindhi-dialect Saraiki ESL learners experienced challenges in the production of English monophthongs. The non-native female speakers produced vowel qualities that were relatively closer to those of the native speakers but experienced difficulty maintaining native-like vowel duration. In contrast, the non-native male speakers generally produced shorter vowel durations than the native speakers, suggesting that they did not consistently maintain the required duration of English monophthongs. Overall, these findings indicate that the acoustic properties of English vowels are influenced not only by gender but also by first-language (L1) interference during second-language (L2) speech production.

Funding: This study was not funded in any shape or form by any party.

Conflict of Interest: The author declares that he has no conflict of interest.

Bio-note:

Abdul Manan is a scholar of linguistics and literature. His research interests include acoustics, sociolinguistics, phonetics, phonology, and second-language acquisition. He has engaged in several research-driven activities and aims to continue his scholarly growth through advanced studies in linguistics. He also studies literature to broaden his critical and analytical perspective.

Wafa Mansoor Buriro is an Assistant Professor and coordinator for Bachelors and Masters programs at Department of English, Sindh Madressatul Islam University Karachi, Sindh. He is currently a doctoral student at the University of Sindh Jamshoro. He has done his MS in Applied Linguistics from NED University of Engineering and Technology, Karachi in 2019 with the highest score among his batch. Earlier, he had completed his BS in Applied Linguistics from University of Sindh, Jamshoro in 2013 with the distinction of first position along with silver medal. He has ten years of teaching experience at university level. Mr. Buriro has authored and/or co-authored 15 research papers which are published in various research journals of national and international level. He has translated three books. His research interest lies in sociolinguistics, language ecology, and lexico-morphology. He has recently finished an HEC-Pakistan funded NRPU project as a Co-PI. He is also the Managing Editor of Sindh Journal of Linguistics (SJL), an HEC recognized and DOAJ-indexed journal.

Muhammad Irfan Khan is an experienced English language educator and linguistics expert with extensive teaching experience in both contemporary and Madarsa settings. He specializes in TESL, expository writing, and curriculum development, with a focus on student engagement and motivation. Currently pursuing an MS in English Linguistics, he combines innovative teaching methods with expertise in media literacy and online education.

References

Aalto, D., Malinen, J., & Vainio, M. (2018). Formants. In *Oxford Research Encyclopedia of Linguistics*.

Abbasi, A. M., Arain, A. A., Khokhar, S., & Batool, M. (2022). An instrumental study of speech variability of English back-rounded vowels between English and Karachiites' Pinglish: Acoustics of English back-rounded vowels. *Pakistan Journal of Languages and Translation Studies, 10*(2), 40–60.

Abbasi, A. M., Husain, I., Bakhsh, I., Khosh, N. K., & Khan, A. (2025). Investigating voice onset time in Pakistani English speech. *Speech Communication, 168*, 103174.

Abbasi, A. M., Iqbal, M. J., & Lakhani, H. Y. (2021). An acoustic analysis of vowel sounds in Kachchi Sindhi. *University of Chitral Journal of Linguistics & Literature, 5*(2), 374–393.

Alam, M., & Khan, A. (2015). The influence of Saraiki on the English vowel production of Saraiki speakers. *Journal of Second Language Pronunciation, 1*(2), 105–123.

Ali, E. M. T. (2013). Pronunciation problems: Acoustic analysis of the English vowels produced by Sudanese learners of English. *International Journal of English and Literature, 4*(10), 495–507.

Atta, F. (2019). *Phonetics and phonology of the Saraiki language: A descriptive exploration and an analysis from the perspective of Optimality Theory* (Doctoral dissertation, Shanghai International Studies University).

Demirezen, M. (2020). Long vowels versus diphthongs in North American English: Which one is easy to recognize and hear? *International Journal of Curriculum and Instruction, 12*(1), 289–309.

Derakhshan, A., & Karimi, E. (2015). The interference of first language and second language acquisition. *Theory and Practice in Language Studies, 5*(10), 2112–2117.

Di Benedetto, M. G. (1994). Acoustic and perceptual evidence of a complex relation between F1 and F0 in determining vowel height. *Journal of Phonetics, 22*(3), 205–224.

Fiaz, H. M., Rind, A. A., Gul, W. A., Akhtar, S., Naseem, H., & Noor, D. (2022). The origin and evolution of Saraiki culture and civilization in South Asia: A case of Saraiki culture from the ancient Indus Valley Civilization to the modern era. *Journal of Positive School Psychology, 6*(7), 5934–5945.

Haan, J., & Van Heuven, V. J. (1999, August). *Male vs. female pitch range in Dutch questions*. Paper presented at the Linguistic Department Conference.

Hansen Edwards, J., Chan, K. L. R., Lam, T., & Wang, Q. (2021). Social factors and the teaching of pronunciation: What the research tells us. *RELC Journal, 52*(1), 35–47.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5), 3099–3111.

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, 9*(1), 353–365.

Joyo, M. A., & Memon, R. A. (2023). Analysis of English phonological performance between Sindhi male and female learners. *Pakistan Languages and Humanities Review, 7*(4), 330–339.

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, 8*(1), 850–863.

Khan, A., Shehzad, W., & Ullah, I. (2017). Articulation of English consonants, vowels, and diphthongs by Pashto speakers in Khyber Pakhtunkhwa, Pakistan. *International Journal of English Linguistics, 7*(5), 19–33.

Khan, M., & Rashid, A. (2010). An overview of the Saraiki language. *Language and Linguistics, 11*(1), 67–82.

Markova, D., Richer, L., Pangelinan, M., Schwartz, D. H., Leonard, G., Perron, M., ... & Paus, T. (2016). Age- and sex-related variations in vocal-tract morphology and voice acoustics during adolescence. *Hormones and Behavior, 81*, 84–96.

Moore, R. K., & Skidmore, L. (2019). On the use/misuse of the term *phoneme*. *arXiv*. https://arxiv.org/abs/1907.11640

Muneer, S. (2021). Women empowerment and access to education in Pakistan: Barriers within the home. *Journal of the Research Society of Pakistan, 58*(2), 32–41.

Pratiwi, D. R., & Indrayani, L. M. (2021). Pronunciation error on English diphthongs made by EFL students. *Teknosastik, 19*(1), 24–30.

Pýcha, J. (2016). *Common pronunciation mistakes among Spanish speakers of English with the focus on vowels*.

Roach, P. (2001). *Phonetics*. Oxford University Press.

Roach, P. (2009). *English phonetics and phonology* (4th ed.). Cambridge University Press.

Roach, P. (2024). *English phonetics and phonology*. Oxford University Press.

Safeer, N., Anjum, U., & Saleem, T. (2024). Gender-based study of paired monophthongs: A sociophonetics approach. *3L: Language, Linguistics, Literature, 30*(2), 231–262.

Shackle, C. (2003). Punjabi. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages* (pp. 581–621). Routledge.

Smith, J. (2011). The use of visual aids in the teaching of English monophthongs to Saraiki speakers. *TESOL Journal, 2*(3), 212–223.

Yaqub, M. (2025). Regional and gender-based acoustic variation in Pakistani English vowels: A sociophonetic analysis of thirteen urban and peripheral varieties. *Liberal Journal of Language & Literature Review, 3*(3), 1527–1559.

Zhang, Z. (2021). Contribution of laryngeal size to differences between male and female voice production. *The Journal of the Acoustical Society of America, 150*(6), 4511–4521.

Zsiga, E. C. (2013). *The sounds of language: An introduction to phonetics and phonology*. John Wiley & Sons.

Zue, V., & Cole, R. (1979, April). Experiments on spectrogram reading. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '79)* (Vol. 4, pp. 116–119). IEEE.