Chapter 2: Phoneme Inventory of Chokri

2.1. Introduction

In the previous chapter, we introduced the Chokri language, outlining its linguistic classification and the status of the language. This chapter provides a detailed description of Chokri's segmental inventory. In a previous study on Chokri (Phek dialect), Bielenberg and Nienu (2001) documented a total of 39 consonants and 6 vowels. While their work provides valuable insights, it is essential to note that their study was conducted on a different dialect (Phek dialect) and relied on impressionistic judgments rather than instrumental analysis. This study builds upon and refines the earlier observations by employing experimental and acoustic methodologies to provide a more objective and detailed analysis.

2.2. Chokri Phonemes

Chokri phonemic inventory consists of 34 consonants[1] and 7 vowels. This inventory reflects a systematic analysis of the phonological system and incorporates both primary data elicited directly from native speakers and observational recordings.

2.2.1. Consonants

Chokri consonant inventory includes nine stops, seven nasals, seven fricatives, six affricates, one lateral, two tap, and two approximants. A three-way distinction among voiced, voiceless, and aspirated features is observed in this language.

2.2.1.1. Plosives

Chokri has a total of nine plosive sounds, which are divided into three types: voiced plosives, voiceless unaspirated plosives, and voiceless aspirated plosives. Chokri plosives are produced from three different places of articulation: bilabial, alveolar and velar. The minimal pairs in the table 2.1 below exhibits examples of the 9 plosives and confirm their presence in the language.

Phoneme	Word	Gloss
[p]	[pɛ́]	Mushroom
[b]	[bɛ́]	To wear shawl
[pʰ]	[pʰɛ̄]	To go straight
[t]	[tɛ̄]	To cultivate
[d]	[dɛ̄]	To arrange

[1] The distribution of Chokri consonats will be discussed in chapter 3.

[tʰ]	[tʰɛ̋]	To burn
[k]	[kə̄]	Stitch
[g]	[gə̄]	Melt
[kʰ]	[kʰə̄]	To make bed

Table 2.1: Minimal pairs of Chokri plosives.

The distinction between the three types of stops can be explained by examining Voice Onset Time (VOT). VOT is the "interval between the release of a consonant (usually a stop) and the start of the voicing of the following vowel" (Ladefoged, 2011). Abramson (1977) suggested that VOT is the most reliable acoustic indicator for distinguishing between different voicing categories.

The VOT can vary across languages due to differences in phonetic and phonological settings, influencing how voicing begins in comparable consonants. Voicing can begin before or after the burst, depending on the language. "Voicing detected before the release or during stop occlusion is called the voicing lead, while voicing starting after the release is called voicing lag" (Gope, 2016). Abramson (1977) emphasizes that negative VOT values indicate voicing onset before the release of a stop, while positive values indicate voicing that begins after the release. For example, In Polish, voiced stops exhibit negative VOT, as voicing begins before the stop is released (Keating et al., 1981). In contrast, English voiced stops typically have a near-zero or slightly positive VOT, meaning voicing begins just as or slightly after the stop release (Keating et al., 1981; Ladefoged, 2011).

Similarly, in Chokri, when producing voiced stops, voicing begins before the release of the stop, resulting in a long negative VOT. i.e., the vocal cords start vibrating before the stop is released. The voiceless stops have a short, positive VOT, as the vocal cords vibrate shortly after the stop is released. And in voiceless aspirated stops, voicing starts after a longer delay following the release, allowing for a clear burst of air (aspiration), which produces a long, positive VOT. These differences can be visualized in waveform diagrams showing the onset of voicing for each type in figure 2.1.

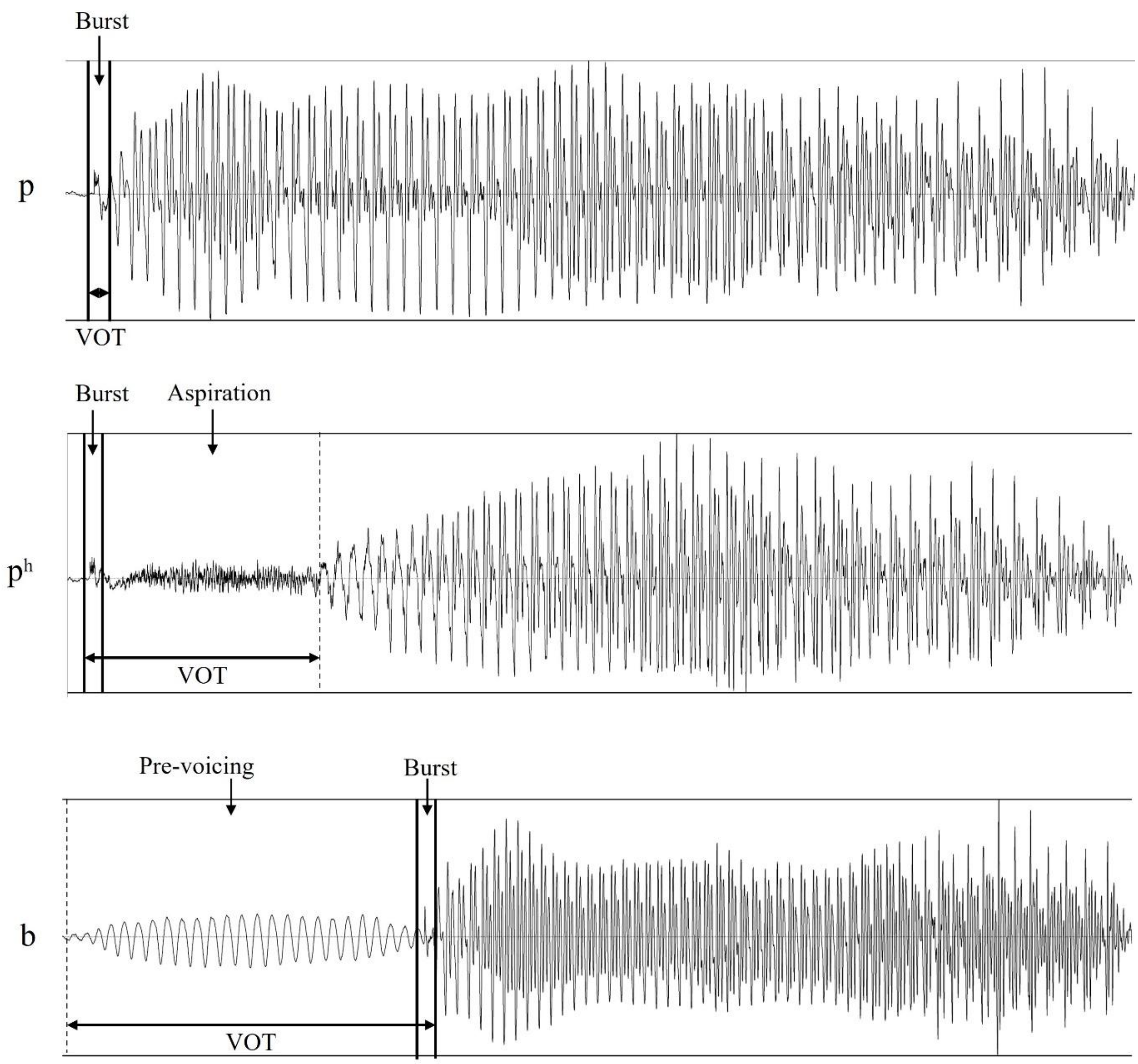

Figure 2.1: Waveforms indicating three types of stops in Chokri in reference to VOT

2.2.1.2. Nasals and Laterals

Chokri has seven nasal sounds, which are divided into two types: voiced nasals and voiceless aspirated nasals. Chokri nasals are produced from four different places of articulation: bilabial, alveolar, palatal and velar. Table 2.2 below shows minimal pairs of nasal sounds found in Chokri. These examples help demonstrate the presence and contrast of nasal sounds in the language.

Phoneme	Word	Gloss
[m]	[mɔ̀]	No
[m̥ʰ]	[m̥ʰɔ́]	Dream
[n]	[nɔ̀]	To breastfeed

[n̥]	[n̥ʰá]	Grass
[ɲ]	[ɲɔ̀]	To run over
[ɲ̥ʰ]	[ɲ̥ʰɔ̀]	To mop
[ŋ]	[ŋɔ̋]	To see

Table 2.2: Minimal pairs of Chokri nasals.

Nasals share an articulatory feature with stops as both stops and nasals involve closure in the oral cavity. However, nasals diverge from stops because they allow airflow through the nasal cavity, resembling approximants due to their relatively open airflow. While most languages produce nasals with modal vocal fold vibration, some languages use alternative phonation types including breathy, creaky, or voiceless quality, to articulate nasal sounds (Ladefoged and Maddieson, 2008).

Our study identified two types of nasals:

(i) Nasals that are produced with typical modal vocal fold vibration and characterized by resonant qualities,

(ii) Nasals that are produced with open glottis lead to the production of voiceless aspirated nasals.

Voiced and voiceless nasals are categories of nasal consonants distinguished by the presence or absence of vocal fold vibration during their articulation. Voiceless nasals (Maddieson, 1984, as cited in Ladefoged and Maddieson, 2008) typically involve an open glottis throughout much of the articulation, with slight voicing occurring just before the closure is released. These sounds tend to be longer than their voiced counterparts and are marked by a higher fundamental frequency (F0) at the beginning of the following vowel.

Voicing contrasts in nasals have been documented in the Tibeto-Burman languages such as: Burmese, Achang, Xumi, Pumi, Zhaba, Chepang, Dhimal, Kham Tibetic, Anong, Angami, Chokri, Khezha, Lushai, Lai, Laizo, Kom Rem etc. (matisoff, 2003). Additionally, Ladefoged and Maddieson, (2008) also reported the same in Burmese and Mizo, where voiceless nasals exhibit partial voicing toward the end of the closure phase. However, unlike Burmese and Mizo, Ladefoged and Maddieson, (2008) described Angami nasals as voiceless aspirated rather than simply voiceless nasals.

Terhiija and Sarmah (2021) conducted a study on Angami, focusing on voiceless aspirated nasals. Their findings indicate that in isolation, Angami voiceless nasals remain entirely voiceless. However, when produced within a sentence frame, a voiced nasal segment may occur at the onset. It was also reported that after the onset of the nasal, oral aspiration begins to appear. Notably, no significant voicing was reported toward the terminal boundary of these voiceless nasals.

The spectrogram provided in Figure 2.3 primarily represents a nasal sound followed by a vowel produced in isolation. Unlike the nasal reported in Angami where voiceless nasal is completely devoid of voicing when produced in isolation. The figure in 2.3 indicates that the nasal sound begins with a brief period of voicing, as indicated by the periodic waveform and the presence of low-frequency energy in the spectrogram. However, following this voiced segment, the waveform loses its periodicity, and the energy in the lower frequencies diminishes significantly, indicating a shift to a fully voiceless phase. This voiceless segment appears as a nearly silent gap in the spectrogram. After this voiceless phase, a faint aspiration phase emerges, characterized by weak and diffuse high-frequency energy. This suggests the presence of an oral airflow component before transitioning into the following vowel. The transition into the vowel is marked by a sudden increase in energy and clear formant structures, indicating the onset of full voicing.

Based on the spectrogram analysis and comparisons to previous studies on voiceless aspirated nasals in Naga languages, the nasal appears to be a voiceless aspirated nasal with voicing at the onset of the nasal. The post-nasal aspiration though faint is present and serves as a distinguishing feature of this sound.

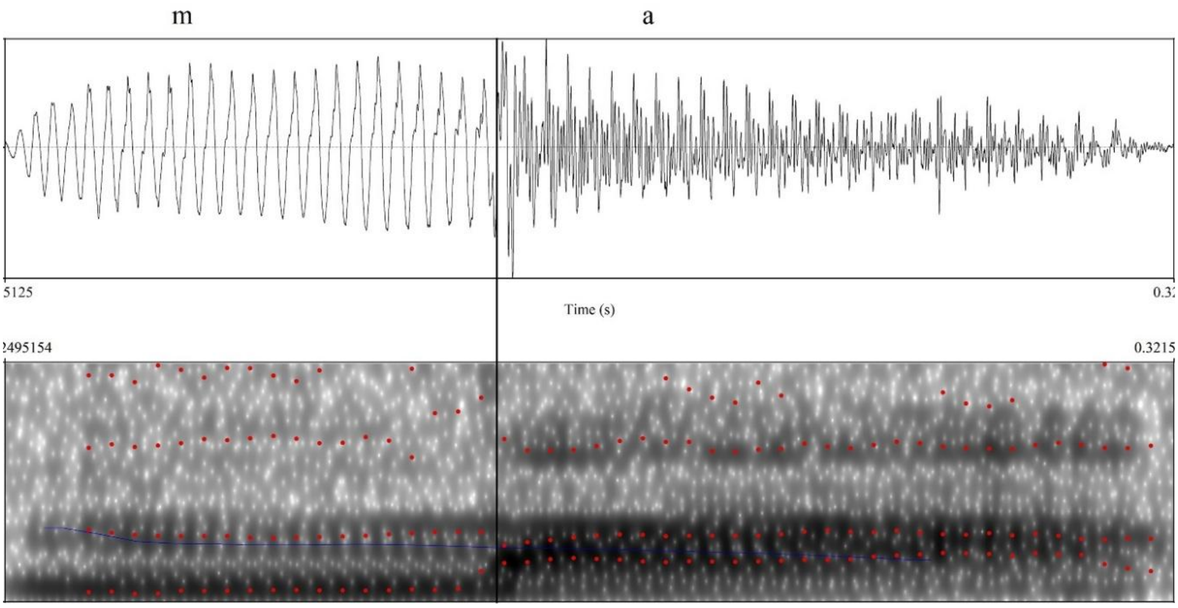

Figure 2.2: Spectrograms illustrating Chokri Voiced bilabial Nasal in the word [ma].

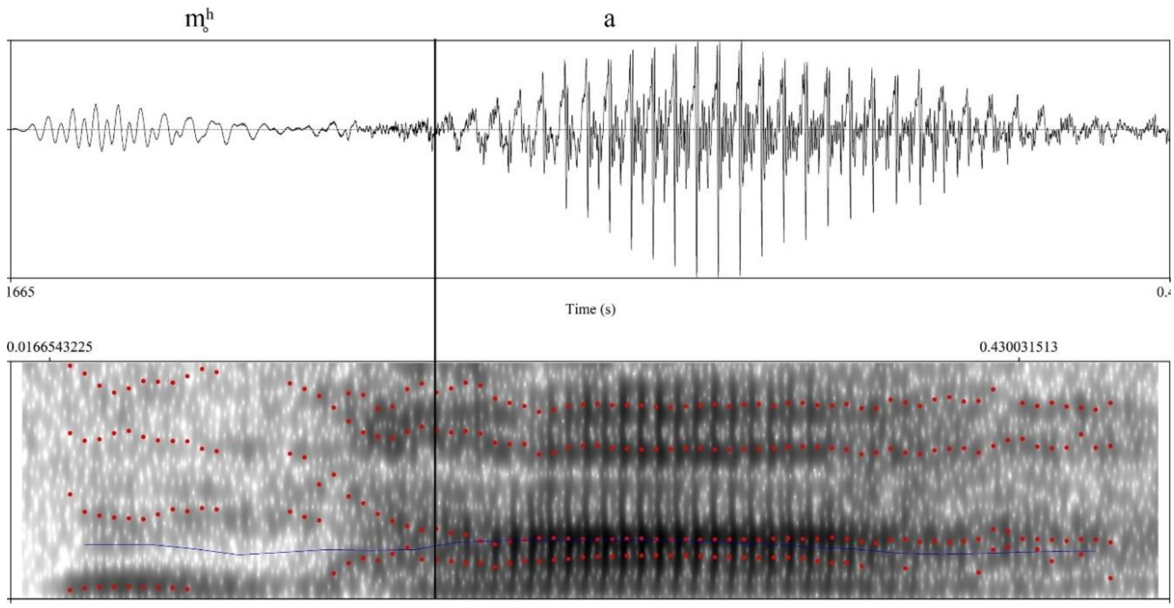

Figure 2.3: Spectrograms illustrating Chokri voiceless aspirated bilabial Nasal in the word [m̥ʰa].

In Sumi, another Naga language, Teo (2014) reports the presence of breathy aspirated nasal which occurs exclusively in syllable-initial position. Chokri voiceless aspirated nasals can occur both word-initially and medially, unlike Sumi, where an epenthetic vowel is inserted in word-medial positions (Teo, 2014). In Chokri, the voiceless aspirated nasals can naturally appear in medial positions without any additional phonetic adjustments.

Word-initial	**Word-medial**
[m̥ʰābə̄] ‘container’	[kə̄m̥ʰā] ‘same’
[n̥ʰātʃɔ́] ‘ant’	[tə̄n̥ʰā̆] ‘algae’
[ɲ̊ʰɔ̀] ‘mop with mud’	[ɛ̄ɲ̊ʰū] ‘caress’

Table 2.3: Chokri voiceless aspirated nasals in initial and medial position.

2.2.1.3. Approximants and Flap

Chokri features two approximants and two taps: the bilabial approximant [w], the palatal approximant [j], and the alveolar taps [ɾ] (voiced) and [ɾ̥](voiceless).

Phoneme	Word	Gloss
[ɾ]	[ɾə̄]	To pour

[ɾ̥]	[ɾ̥ə̄]	To be alive
[w]	[āwú]	We (dual, inclusive)
[j]	[ji]	Field

Table 2.4: Minimal and near minimal pairs of Chokri approximants and taps.

2.2.1.3.1. Bilabial Approximant [w]

The bilabial approximant [w] in Chokri demonstrates characteristics of a marked phoneme due to its limited occurrence and specialized functional role in the language. While phonemic in status, [w] is predominantly found in interjectional words rather than across the broader lexicon, setting it apart from other more pervasive phonemes. The bilabial approximant [w] though occurs infrequently, its phonemic significance is affirmed by its contrastive distribution and its ability to appear in all permissible syllable positions.

2.2.1.3.2. Alveolar Taps [ɾ] and [ɾ̥]

An interesting sound identified in Chokri is the voiceless unaspirated rhotic tap [ɾ̥], which stands out as a typologically rare phenomenon. Voiceless unaspirated rhotics are scarcely documented globally, making their presence in Chokri linguistically significant. This sound, characterized by the quick contact of the tongue against the alveolar ridge combined with voicelessness, contrasts acoustically and articulatorily with the more common [ɾ].

The voiced unaspirtaed rhotic tap [ɾ], as found in many languages such as Spanish (Ladefoged and Maddieson, 2008), exhibits a short closure duration, typically around 20 ms (Quilis, 1981, as cited in Ladefoged and Maddieson, 2008) and distinct formant transitions into and out of the tap. Acoustically, [ɾ] is marked by its low-intensity burst during the release phase and clear voicing throughout the articulation. The energy distribution reflects voicing with a periodic waveform and clear harmonic structure, especially noticeable in lower frequencies. The third formant (F3) remains relatively stable, distinguishing [ɾ] from other rhotics such as approximants: which tends to lower towards the second formant (Ladefoged and Maddieson, 2008).

In contrast to the voiced unaspirtaed rhotic sound, the voiceless rhotic [ɾ̥] in Chokri is articulated with a similar brief contact but with an open glottis, resulting in the absence of vocal fold vibration. Acoustically, [ɾ̥] exhibits a sharper burst compared to [ɾ], and the lack of periodic voicing results in a noisier waveform. Formant transitions are still observable but less prominent, as the voicelessness reduces harmonic clarity. A higher fundamental frequency (f0) at the onset of the following vowel often characterizes voiceless unaspirtaed rhotics, a feature also noted in voiceless nasal sounds.

In Chokri, [ɾ̥] occurs in all permissible positions: initial, medial, and in consonant clusters highlighting its functional and phonemic significance.

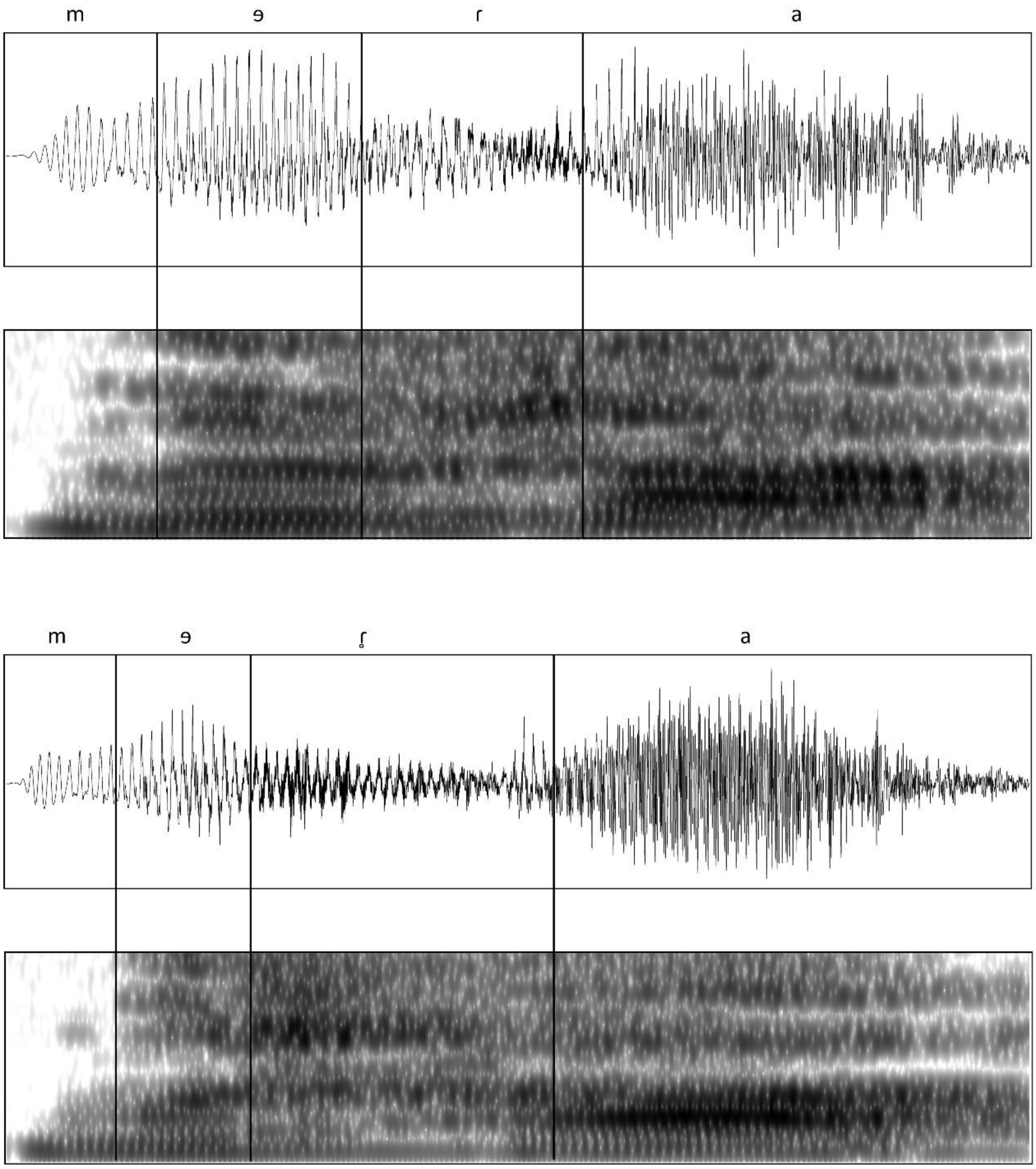

Figure 2.4: Spectrograms of voiced rhotic tap in the word [məɾa] 'bird' and voiceless rhotic tap in the word [məɾ̥a] 'to scratch'

2.2.1.4. Fricatives

Chokri has seven fricative consonants, which can divided into two types based on their voicing property: voiced unaspirated fricatives and voiceless unaspirated fricatives. Chokri fricatives are produced from four different places of articulation: labiodental, alveolar, post-alveolar, and glottal. With the exception of glottal fricatives, all fricative types distinguish between voiced and voicelessness. The table below presents examples of these fricatives, showing the contrast between the two types.

Phoneme	Word	Gloss
[f]	[fə̋]	To wait
[v]	[və̋]	To beat
[s]	[sə́]	To get up
[z]	[zə́]	To lie down
[ʃ]	[ʃə̄]	To wipe
[ʒ]	[ʒə̋]	To bear fruit
[h]	[hə̄]	To live

Table 2.5: Minimal pairs of Chokri fricatives.

Acoustically, voiced unaspirtaed fricatives differ from voiceless unaspirated ones due to the presence of vocal fold vibration, which appears in a spectrogram as low-frequency energy (often below 300 Hz). Voiceless unaspirated fricatives, lacking this voicing, are primarily characterized by higher-frequency noise.

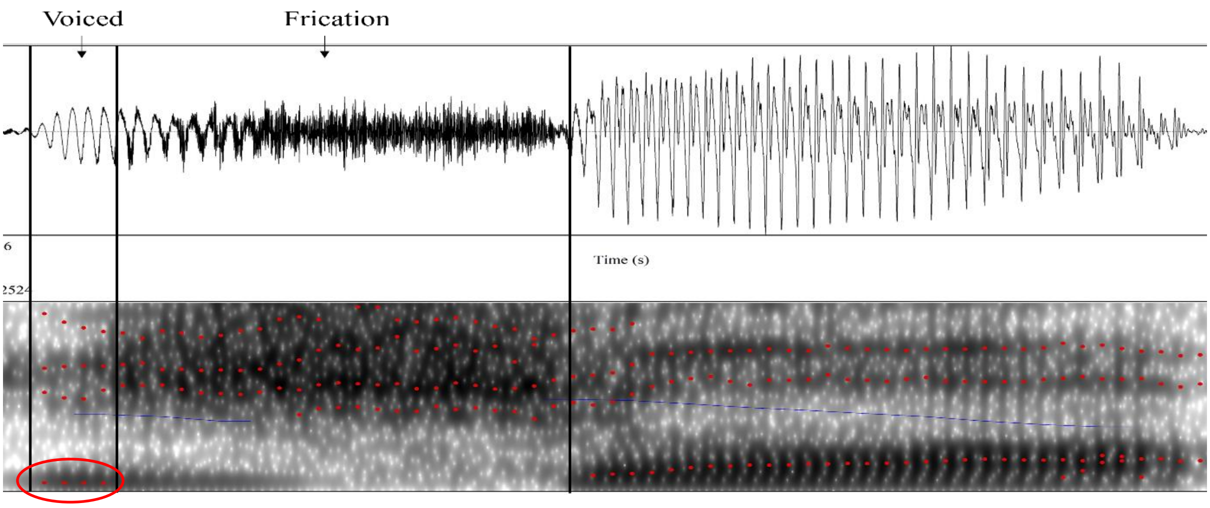

Figure 2.5: Waveform and spectrograms illustrating Chokri voiced unaspirated fricative in the word [və].

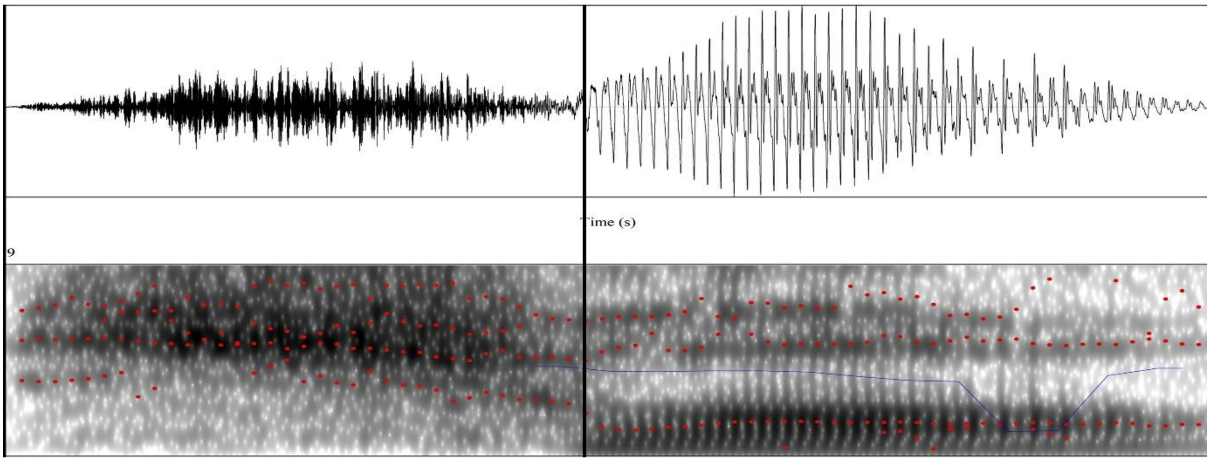

Figure 2.6: Waveform and spectrograms illustrating Chokri voiceless unaspirtaed fricative in the word [fə].

Labiodental and alveolar fricatives exhibit lower spectral peaks than post-alveolar fricatives, which show more energy concentrated at slightly higher frequencies due to the more posterior place of articulation.

2.2.1.5. Affricates

Chokri has six affricate consonants, which are divided into three types: voiced unaspirated, voiceless unaspirated and voiceless aspirated affricates. Chokri affricates are produced at three different places of articulation: labiodental, alveolar, post-alveolar. The table below presents examples of these affricates, showing the contrast between the three types.

Phoneme	Word	Gloss
[ts]	[tsə́]	Small
[tsʰ]	[tsʰə̄]	To sundry
[tʃ]	[tʃə̋]	Call
[dʒ]	[dʒə́]	Water
[tʃʰ]	[tʃʰə́]	To dig
[pf]	[pfə̋]	To collect (snailing)

Table 2.6: Minimal pairs of Chokri affricates.

Affricates are often characterized as an intermediate category between simple stops and a sequence of a stop and a fricative (Ladefoged and Maddieson, 2008). Acoustically, this structure creates a burst of energy similar to a stop, followed by sustained noise from the fricative element. These distinctive patterns of their stops and fricative releases across voiced, voiceless, and voiceless aspirated affricates can be observed in the waveforms (Figures 2.7-2.9).

(i) Voiced Unaspirtaed Affricates: These sounds have a voiced stop closure, which can be identified by periodic voicing in the waveform and spectrogram. The release phase is less abrupt and contains lower amplitude frication noise due to continuous vocal fold vibration.

(ii) Voiceless Unaspirtaed Affricates: These are characterized by a voiceless stop phase with a sudden burst release followed by high-frequency friction noise. On the spectrogram, the voiceless stop appears as a gap, followed by a fricative section with visible energy in higher frequencies.

(iii) Voiceless Aspirated Affricates: These sounds feature a voiceless stop with an added period of aspiration before the fricative. In the spectrogram, aspiration shows up as low-amplitude noise immediately after the release, which then transitions into a high-frequency fricative noise.

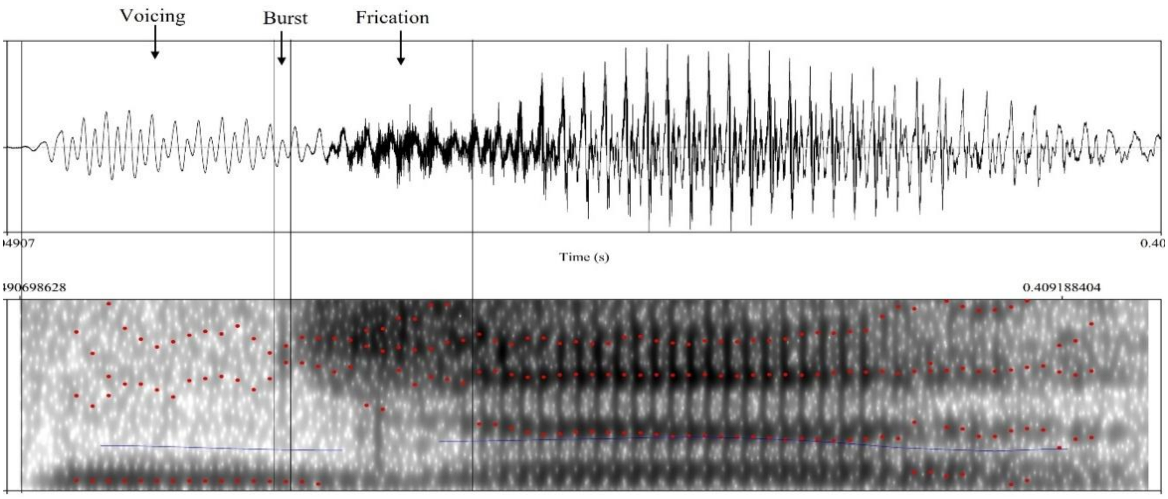

Figure 2.7: Waveform and spectrograms illustrating voiced unaspirated affricate in the word [dʒə].

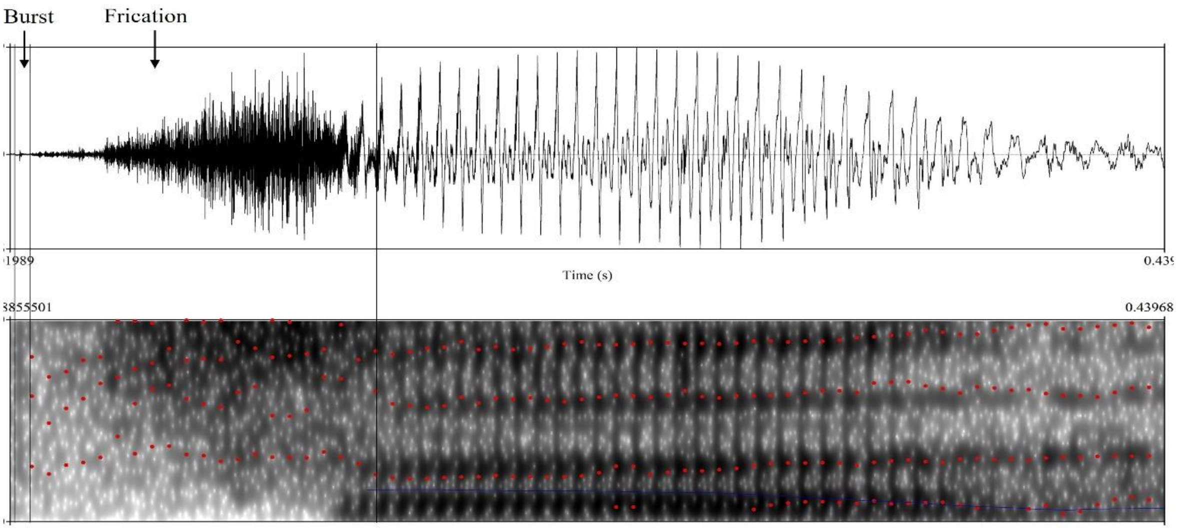

Figure 2.8: Waveform and spectrograms illustrating voiceless unaspirated affricate in the word [tʃə].

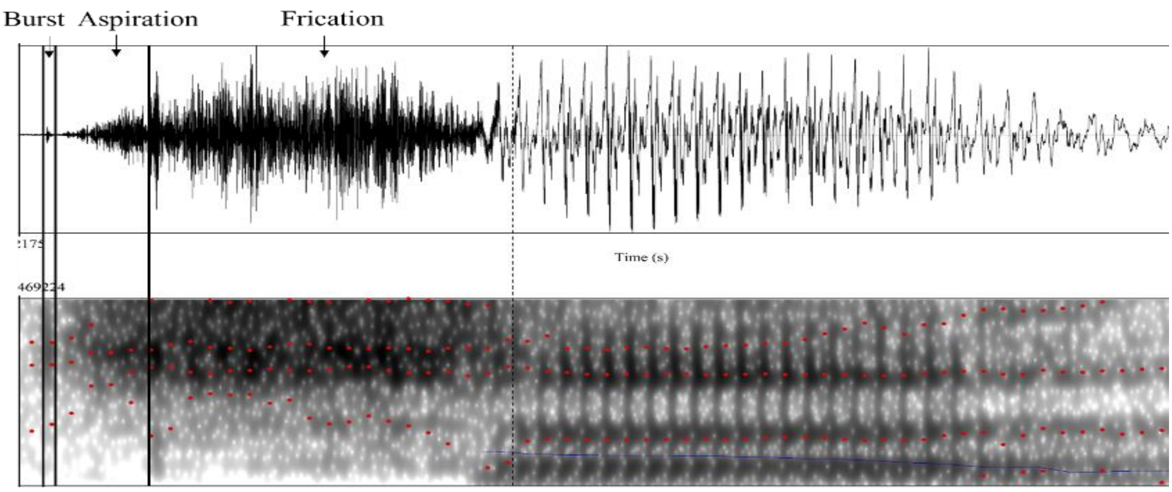

Figure 2.9: Waveform and spectrograms illustrating voiceless aspirated affricate [tʃʰə].

	Bilabial	*Labio-dental*	*Alveolar*	*Post-alveolar*	*Palatal*	*Velar*	*Glottal*
Plosive	p b		t d			k g	
	p^h		t^h			k^h	
Nasal	m m̥^{h}		n n̥^{h}		ɲ ɲ̊^{h}	ŋ	
Fricative		f v	s z	ʃ ʒ			h
Affricate		pf	Ts	tʃ dʒ			
			ts^h	tʃ^h			
Lateral			l				
Tap/flap			ɾ̥ ɾ				
Approximant	w				j		

Table 2.7: Consonants observed in Chokri.

2.2.2. Vowel

Chokri has a distinctive vowel system that includes seven primary vowel sounds. The table below presents the minimal pairs that would establish the vowel inventory in Chokri.

Vowels	Vd_vowel	Gloss	Vl_Vowel	Gloss
i	[bi]	Taro	[pi]	Head
ɛ	[bɛ]	Hand	[pɛ]	Look
a	[ba]	Sit	[pa]	bald (v)
ɔ	[bɔ]	Surround	[pɔ]	Drip
u	[bu]	to spread out cloth to collect	[pu]	masculine marker
ɘ	[bɘ]	To trim/cut	[pɘ]	Take
ə	[ʙə][2]	To defecate	[ʙ̥ə]	to carry on back

Table 2.8: Minimal pairs observed in Chokri.

All vowels in Chokri, except for the mid-central vowels [ɘ] and [ə], can occur in words' initial medial and final positions. Vowels [ɘ] and [ə] are only found in the medial and final positions. The distribution of Chokri vowels is shown in the table 2.9.

[2] [ʙ] and [ʙ̥] are allophones of the phoneme [b] and [p] respectively. They occur with the central vowel [ə].

Vowel	Initial	Gloss	Medial	Gloss	Final	Gloss
i	[**i**]	1sg	[tʰɛz**i**bɘ]	Bedroom	[satʰ**i**]	Earthworm
ɛ	[**ɛ**rɔ]	Rope	[pʰ**ɛ**ka]	Heel	[mɛl**ɛ**]	Climb
a	[**a**-bɛ]	My hand	[tʰ**a**ɾɛ]	Splits end	[ɾəd**a**]	Leech
ɔ	[**ɔ**hɛ]	Rhino	[pʰ**ɔ**tɘ]	Whole body	[sɘd**ɔ**]	tomorrow
u	[**u**-kɔ]	We	[ʃ**u**jɛ]	Generation	[mɛd**u**]	Talk
ɘ	-	-	[mɛs**ɘ**dɔ]	Firewood	[s**ɘ**]	Three
ə	-	-	[tɛʃ**ə**nɛ]	Puppy	[f**ə**]	Fish

Table 2.9: Distribution of vowels in Chokri

2.2.2.1. Production Experiment: Stimuli, Subjects, and Recording Procedure

For the production experiment, the same minimal pairs (Table 2.8) that include each vowel in two different contexts, [Vd_V] and [Vl_V] (where Vd represents a voiced consonant, Vl a voiceless consonant, and V the target vowel) were considered. The study involved five native speakers (2 male and 3 female), aged between 18 and 50 years from Thipüzu village, Phek district, who speak this village dialect as their mother tongue.

Three different steps were employed in recording the speech data.

1. Word with target vowel uttered in isolation.
2. Target words in simple priming sentences, which were intended to guide the subjects at producing the correct word.
3. Target word in fixed sentence frame/ carrier phrase “Say X again”, “Vapü X si sa sü te” X being the word with target vowel.

All three steps were repeated 5 times, yielding a total of 1050 tokens. However, only the tokens produced in isolation and fixed carrier frame were considered for the final analysis. Thus, a total of 700 tokens (14 words x two contexts x five subjects x five repetitions) are examined to understand the properties of Chokri vowels.

The speech data was recorded using a Multi-Track Linear PCM Recorder (OLYMPUS, LS-100) and SHURE headset microphone (SM10) at a sample rate of 44100 Hz in .wav format. The recorded data was then manually segmented and annotated using PRAAT (Boersma and Weenick, 2013). Data pre-processing, data visualization, and statistical analyses were carried out using *R* and *Python*.

2.2.2.2. Acoustic Analysis of Chokri Vowels

The acoustic properties of Chokri vowels were analyzed to examine their spectral characteristics and possible phonetic conditioning effects. The analysis was conducted using Praat, where the starting and ending pitch points of each vowel token were manually labeled. Labels were placed at zero crossings, marking the transition from the onset consonant to the vowel while ensuring that no residual noise from the surrounding segments interfered with the vowel signal. This manual labeling process ensured precision in vowel segmentation, which is crucial for accurate formant extraction. To minimize the influence of consonantal coarticulation, the first three formant frequencies (F1, F2, F3) were measured at the midpoint of each vowel. This approach accounts for the steady-state portion of the vowel, reducing possible transitional effects from the preceding consonant.

The key acoustic indicators for vowels are their formant frequency values, which represent the resonances of the vocal tract (Fant, 1960). A formant is a characteristic resonance of the vocal tract that shapes the acoustic properties of speech sounds. Rosner and Pickering (1994) highlighted the importance of formant values in distinguishing vowel contrasts in cross-linguistic contexts. The first formant (F1) is inversely related to vowel height: lower F1 values correspond to higher vowels, while higher F1 values indicate lower vowels. The second formant (F2) reflects vowel frontness, with higher F2 values corresponding to front vowels and lower F2 values indicating back vowels. Meanwhile, the third formant (F3) is often associated with lip rounding, where specific configurations of the lips influence its frequency (Ladefoged, 2011; Anderson, 2002; Hick, 2017).

Although vowels typically exhibit more than three formants, the higher formants (beyond F3) are not critical for determining vowel quality. Instead, these formants are more closely linked to the speaker's individual voice quality (O'Connor, 1973). Research has consistently shown that the first two formants, F1 and F2, are the most crucial for vowel perception, as they primarily encode vowel height and frontness. For example, Hillenbrand et al. (1995) found that the F1-F2 relationship provides sufficient information for accurate vowel perception in American English. Similarly, a study by Peterson and Barney (1952) established a direct correlation between vowel formant frequencies and their perceptual categories, revealing that F1 and F2 are the primary acoustic cues for distinguishing vowels in American English.

Carlson et al. (1970) showed that the F1-F2 plane could effectively map vowels in Swedish, demonstrating its universality in vowel perception across different languages. Likewise, Rosner and Pickering (1994) provided evidence that the F1-F2 relationship plays a significant role in distinguishing vowels in various Indo-European languages, highlighting its cross-linguistic relevance. Nearey and Assmann (1986) further explored the importance of F1 and F2. Their study reveals that even when vowel tokens were synthesized to exclude higher formants (F3 and beyond), listeners could still accurately identify vowels based on the F1-F2 values. Strange et al. (1998) examined Japanese and American English vowels, showing that the perceptual importance of F1 and F2 extends across linguistic boundaries, even in languages with markedly different vowel inventories.

Consequently, this study analyses F1 and F2 values to define the vowel inventory of Chokri. The acoustic properties of vowels will be detailed based on their height and frontness by examining these primary formants. Additionally, F1 and F2 will play a key role in identifying and mapping the unique vowel qualities, facilitating the construction of a vowel space diagram that illustrates the relative positions of vowels. The analysis of F1-F2 relationships will also allow for cross-gender comparisons. The measured formant values in Hz were transformed into the Mel scale in *R*, which better approximates human auditory perception. To further account for inter-speaker variation, the extracted formant values were normalized using the Lobanov method. Lobanov normalization standardizes vowel formant frequencies by z-score transformation, ensuring that speaker-dependent differences in vocal tract size do not skew the analysis. This acoustic analysis provides a reliable basis for examining vowel quality in Chokri

2.2.2.3. Results and Discussion

Our analysis confirmed the presence of seven vowel phonemes in Chokri. Figures 2.10 and 2.11 present the distribution of the first three formants (raw values measured in Hz), viz., F1, F2, and F3 values for the seven Chokri vowels uttered in two environments, viz., [Vd_V] and [Vl_V], where Vd represents a voiced consonant, Vl a voiceless consonant, and V the target vowel. The horizontal bars within each panel indicate the measured frequency values for each formant (in Hz) across time (in milliseconds). Each panel corresponds to a specific vowel, as indicated by the labels at the top, viz., [i, ɛ, a, ɘ, ə, ɔ, and u].

Figures 2.10 and 2.11 clearly indicate that all the 7 vowels observed in Chokri are distinctly different interms of their first three formant values. As stated above, the raw formant values were converted in Mel to examine the perceptual differences across vowels, and the first two formant values were averaged across all the repetitions by all the male and female speakers separately. These values are used to draw the vowel periphery of Chokri vowels produced by all male and female speakers for visual inspection and to determine the possible vowel quality in Chokri (Figure 2.12).

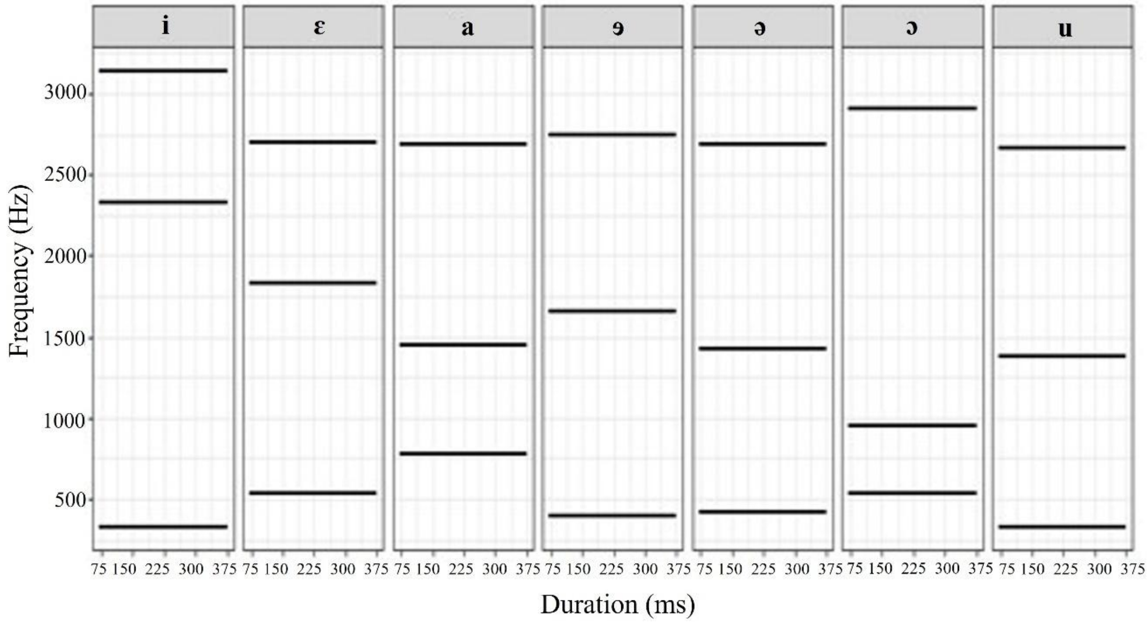

Figure 2.10: Distribution of first three formants (measured in Hz): F1, F2, and F3 for each vowel uttered in the Vl_V environment. Vl= Voiceless consonant occurring in the onset position, V= the target vowel.

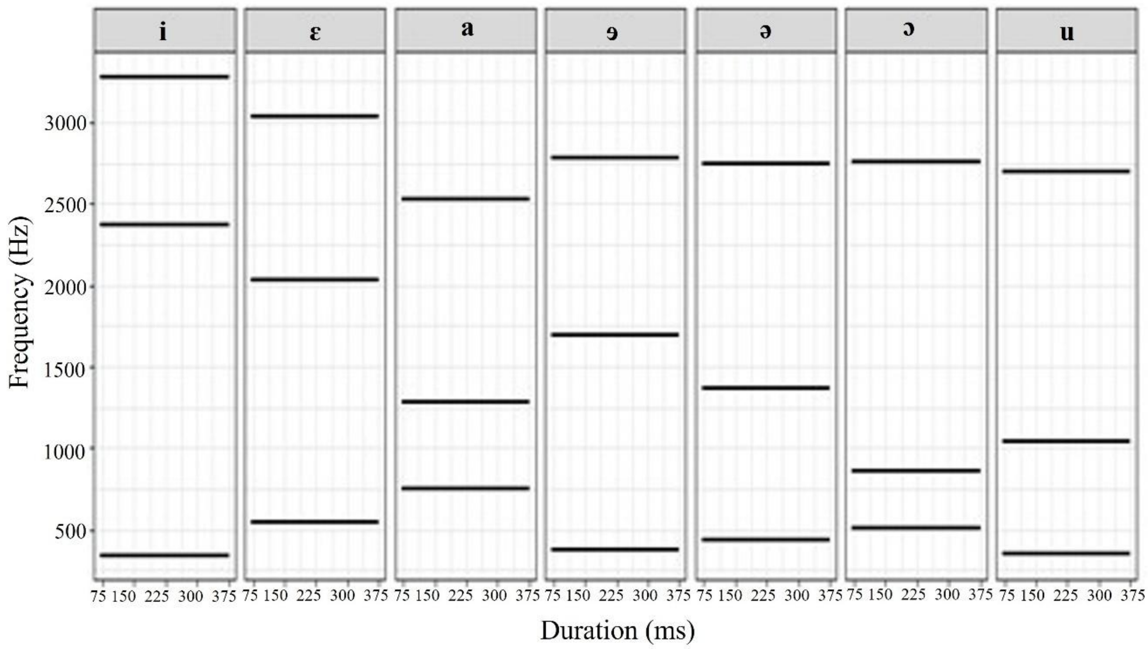

Figure 2.11: Distribution of first three formants (measured in Hz): F1, F2, and F3 for each vowel uttered in the Vd_V environment. Vd Voiceed consonant occurring in the onset position, V= the target vowel.

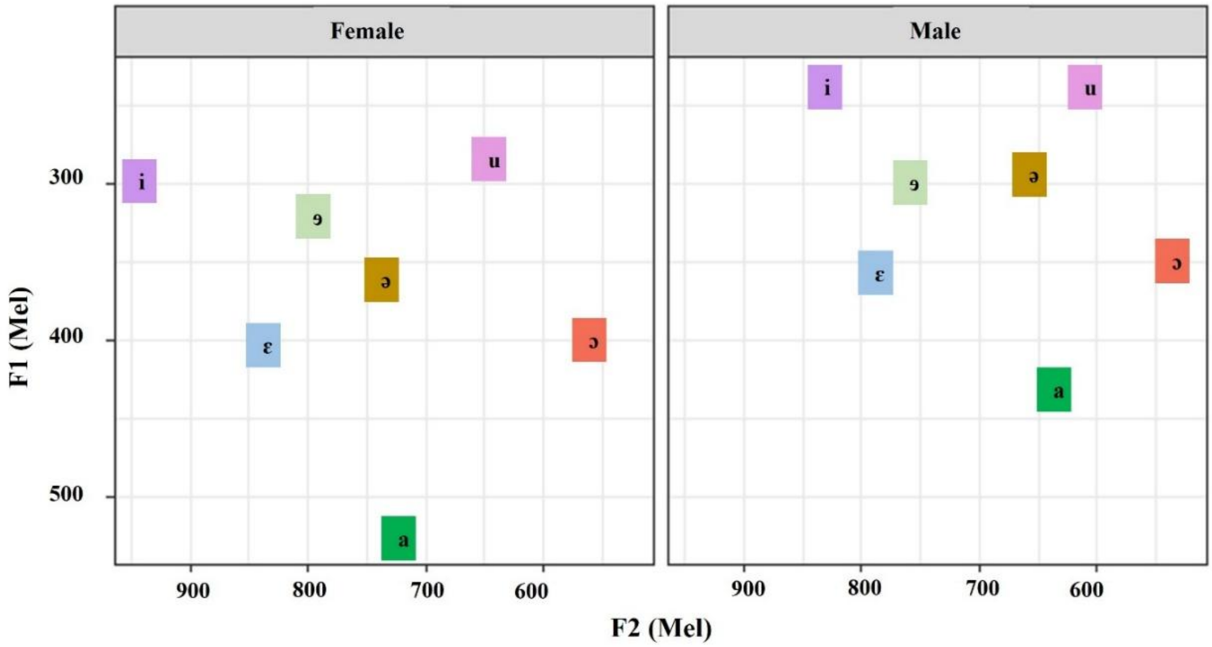

Figure 2.12: Comparison of vowel periphery by female (left panel) and male (right panel) speakers.

Figure 2.12 provides a clear comparison of the vowel periphery of Ckori vowels produced by male and female speakers. This figure also illustrates the possible position of individual vowels in Chokri. We can visualize that there are two front vowels [i and ɛ], three central vowels [a, ɘ, and ə], and two back vowels [ɔ and u]. Interestingly, the low vowel [a] is almost centralized in both male and female speakers' data. This figure also indicates that the vowel space is much narrower in the male speakers' data compared to their female counterparts.

Figure 2.13 shows the overall vowel periphery in Chokri, drawn using the first two formant values in the Y and X axes, respectively. The formant values are averaged separately across all the tokens produced by all the subjects for each vowel.

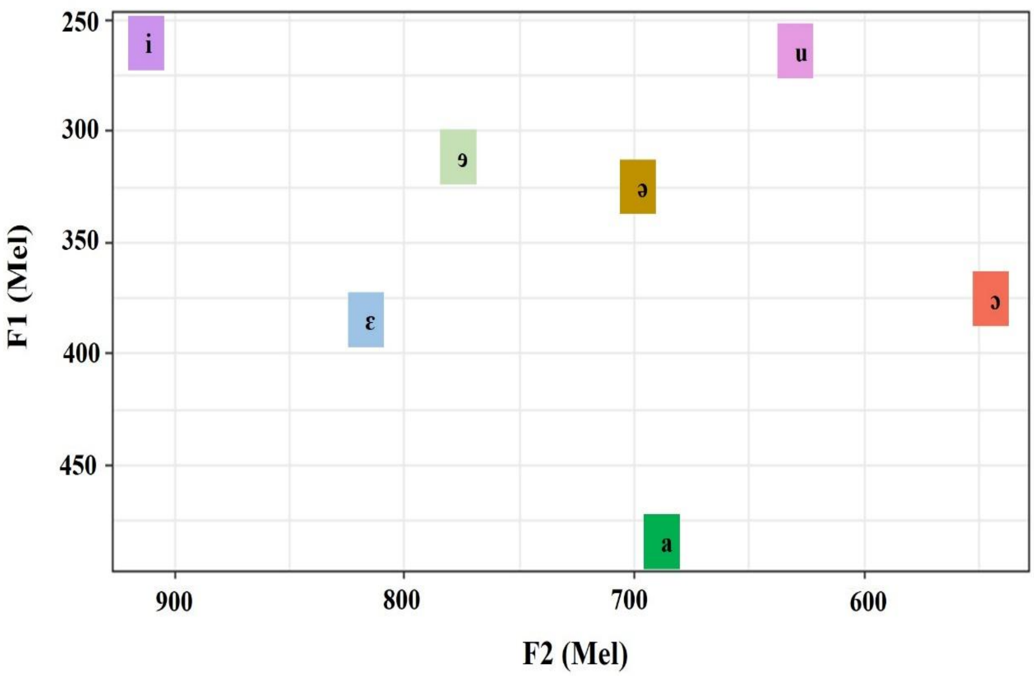

Figure 2.13: Vowel periphery in Chokri, drawn using the first two formant values (in Mel), averaged across all the tokens produced by all the subjects.

Fant (1966) observed that the length and shape of the vocal tract differ significantly between males and females, with males typically exhibiting formant frequencies approximately 20% lower than females due to their longer vocal tracts. According to Bordon and Harris (1980), females' pharynx is approximately 2 cm shorter, and the oral cavity is about 1.25 cm shorter compared to males. These anatomical differences directly influence the acoustic properties of speech sounds.

These differences emphasize the necessity of normalizing formant values when establishing inventory in phonetic research. To address this variability, Lobanov's (1971) normalization procedure is employed to eliminate the effects of individual and gender anatomical differences. Disner (1980) emphasizes that normalization is an effective tool for vowel classification, as it removes speaker-specific effects from the data. In this study, the first two formant values were normalized to account for speaker-specific anatomical differences. This process isolates phonetic and anatomical details by reducing speaker variability, ensuring that the vowel space is comparable across speakers with different vocal tract sizes or genders. Figure 2.14 shows the normalized vowel space in Chokri, drawn using the normalized mean F1 and F2 values, averaged across all the tokens by all the speakers.

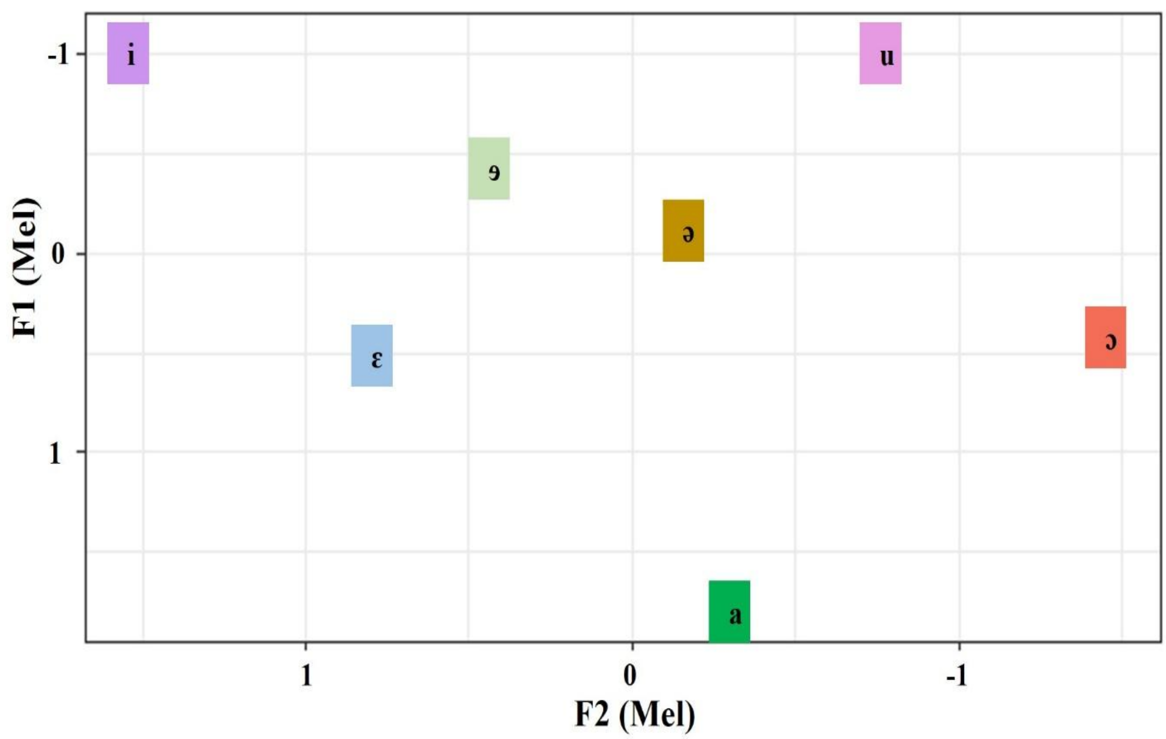

Figure 2.14: Vowel space in Chokri, drawn using the normalized F1 and F2 values, averaged across all the tokens produced by all the subjects for each vowel.

2.2.2.1. Vowel Space Comparison of Male and Female Speakers

Figure 2.15 illustrates the vowel space comparison between male and female speakers in Chokri based on the raw F1 and F2 values (measured in Hz). The vertical axis represents the first formant (F1), and the horizontal axis represents the second formant (F2).

The vowel space diagram provides insights into the articulatory and acoustic properties of vowels produced by male and female Chokri speakers. This figure indicates that the F1 and F2 values for female speakers are consistently higher than those for males across all vowels. Female speakers' vowels (red) occupy a higher and larger range of F1 and F2 values due to their shorter vocal tracts, resulting in relatively higher resonances. Meanwhile, male speakers' vowels (black) show a concentration of formants shifted lower along both the F1 and F2 axes. This suggests that female speakers produce vowels with a broader range of tongue height and advancement. Female vowels exhibit greater dispersion, particularly in high front vowels [i] and low central vowels [a], which occupy more extreme positions. This greater dispersion suggests a more distinct articulation of vowels in females.

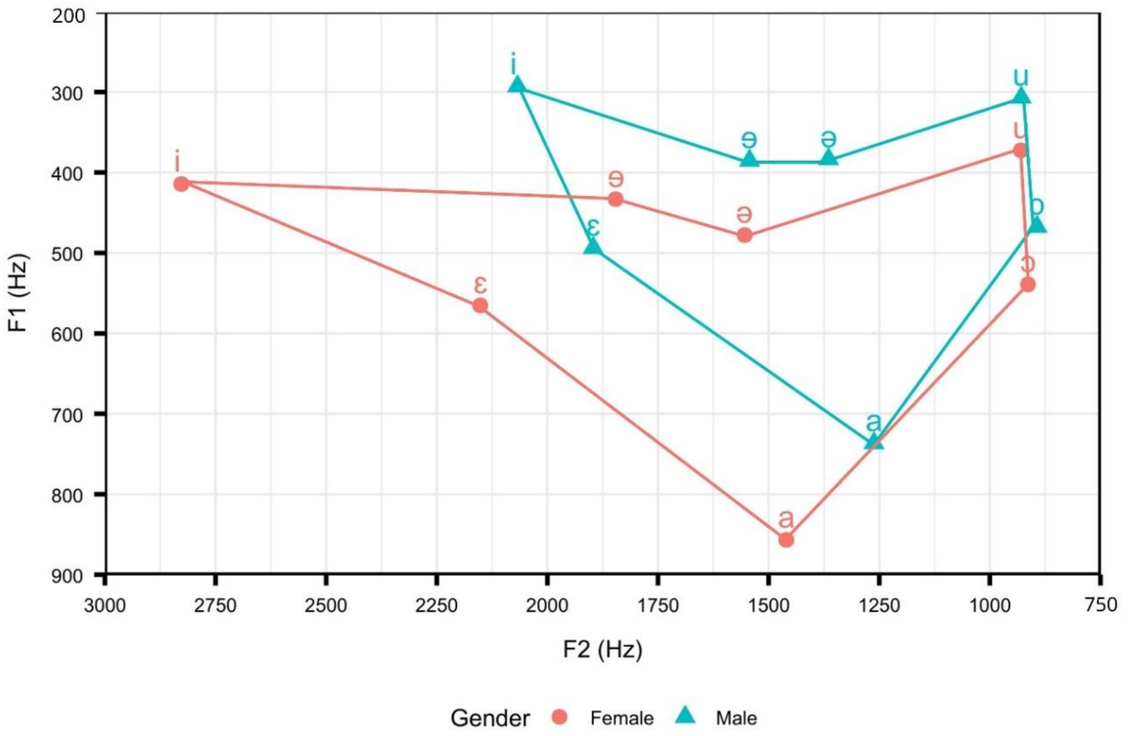

Figure 2.15: Comparison of vowel space in Male and Female speakers.

These non-uniform differences in vowel production between male and female speakers follow the general tendencies across languages such as English (Hillenbrand et al., 1995, as cited in Simpson, 2009), Swedish (Fant, 1966), Hakka (Man, 2007), etc. The difference in the ranges of F1 and F2 values between male and female speakers is directly attributable to the anatomical distinctions discussed by Fant (1966). The formant frequencies of vowels are affected by the length of the pharyngeal-oral tract, the vocal tract constriction, and the degree of narrowness of the constriction (Pickett, 1996).

2.3 Statistical Modeling

The present study employs Generalized Additive Models (GAMs) to analyse the relationship between vowels and three acoustic parameters: formants (F1, F2, and F3), and duration. The dataset was processed using R. The vowels were mapped as a, i, u, ɛ, ɔ, ɘ and ə. The GAM models were computed using *the mgcv* package, and were specified as F1 ~ vowel, F2 ~ vowel, F3 ~ vowel, and Duration ~ vowel. Each model was fitted using a Gaussian family with an identity link function. The estimated marginal means (emmeans) and pairwise comparisons were calculated using the *emmeans* package, adjusting p-values using the Bonferroni method.

It must be noted that we used the raw formant values (measured in Hz) for the statistical tests and modeling.

The analysis of formant values using Generalized Additive Models (GAM) revealed significant differences in the first (F1) and second (F2) formant frequencies across vowel categories. The third formant (F3) exhibited moderate variation, while vowel duration showed no substantial differences. For F1, the vowel [a] has an estimated value of 771 Hz. Other vowels show significantly lower F1 values, consistent with their tongue height. High vowels such as [i and u] exhibit the most substantial reduction in F1, with estimates of -422 Hz and -426 Hz, respectively ($p < 0.001$). Mid vowels, such as [ɛ and ə], also exhibit significantly lower F1 values compared to a. The statistical comparisons confirm that all pairwise differences are significant ($p < 0.001$), except between a few mid-vowel pairs where the distinction is less pronounced. The adjusted R^2 value of 0.769 suggests that the vowel categories explain approximately 77% of the variance in F1, highlighting the strong influence of vowel articulation. For F2, the vowel [a] has an estimated value of 1428 Hz. Front vowels [i and ɛ] exhibit significantly higher F2 values of 988 Hz and 596 Hz, respectively ($p < 0.001$).

In contrast, back vowels [u and ɔ] show lower F2 values. The statistical analysis indicates strong distinctions between front and back vowels, with most pairwise differences reaching significance ($p < 0.001$). The adjusted R^2 value of 0.679 suggests that the vowel categories explain approximately 68% of the variance in F2. For F3, the vowel [a] has an estimated value of 2779 Hz. Although some vowels, such as [i and ɛ], show higher values, the overall variation is less pronounced than for F1 and F2. The adjusted R^2 value of 0.244 indicates that the vowel categories explain only 24.8% of the variance in F3, suggesting additional factors influence this formant.

Unlike the formant frequencies, vowels do not appear to have a significant effect on duration. None of the pairwise comparisons between vowels reached statistical significance at $p < 0.05$. The pairwise comparisons of F1, F2, F3, and duration are given in Tables 2.10-2.13, respectively.

Contrast	**Estimate**	**SE**	**t-value**	**p-value**
a - ɛ	251.45	8.18	30.73	<0.001*
a - i	423.0	8.18	52.45	<0.001*
a - ɔ	258.95	8.18	31.75	<0.001*

a - u	426.66	8.18	52.31	<0.001*
a - ə	335.85	8.18	40.48	<0.001*
a - ɘ	362.87	8.18	44.15	<0.001*
ɛ - i	171.54	8.18	21.18	<0.001*
ɛ - ɔ	7.5	8.18	0.92	1.0000
ɛ - u	175.21	8.18	21.38	<0.001*
ɛ - ə	84.39	8.18	10.13	<0.001*
ɛ - ɘ	111.42	8.18	13.5	<0.001*
i - ɔ	-164.04	8.18	-20.31	<0.001*
i - u	3.67	8.18	0.45	1.0000
i - ə	-87.15	8.18	-10.6	<0.001*
i - ɘ	-60.12	8.18	-7.39	<0.001*
ɔ - u	167.71	8.18	20.53	<0.001*
ɔ - ə	76.9	8.18	9.26	<0.001*
ɔ - ɘ	103.92	8.18	12.63	<0.001*
u - ə	-90.81	8.18	-10.93	<0.001*
u - ɘ	-63.79	8.18	-7.75	<0.001*
ə - ɘ	27.02	8.18	3.23	0.0268*

Table 2.10: Pairwise Comparison for F1 across vowel categories (a, i, u, ɛ, ɔ, ɘ and ə). The * shows the significant pairs.

Contrast	**Estimate**	**SE**	**t-value**	**p-value**
a - ɛ	-596.08	35.33	-16.87	<0.001*
a - i	-988.68	35.33	-28.39	<0.001*
a - ɔ	488.37	35.33	13.86	<0.001*
a - u	263.17	35.33	7.47	<0.001*
a - ə	52.55	35.33	1.47	0.143
a - ɘ	276.68	35.33	7.8	<0.001*
ɛ - i	-392.6	35.33	-11.22	<0.001*

ɛ - ɔ	1084.5	35.33	30.65	<0.001*
ɛ - u	859.2	35.33	24.28	<0.001*
ɛ - ə	543.5	35.33	15.1	<0.001*
ɛ - ɘ	319.4	35.33	8.96	<0.001*
i - ɔ	1477.1	35.33	42.34	<0.001*
i - u	1251.8	35.33	35.89	<0.001*
i - ə	936.1	35.33	26.37	<0.001*
i - ɘ	712.0	35.33	20.25	<0.001*
ɔ - u	-225.2	35.33	-6.38	<0.001*
ɔ - ə	-540.9	35.33	-15.07	<0.001*
ɔ - ɘ	-765.1	35.33	-21.52	<0.001*
u - ə	-315.7	35.33	-8.8	<0.001*
u - ɘ	-539.9	35.33	-15.19	<0.001*
ə - ɘ	-224.1	35.33	-6.2	<0.001*

Table 2.11: Pairwise Comparison for F2 across vowel categories (a, i, u, ɛ, ɔ, ɘ and ə). The * shows the significant pairs.

Contrast	Estimate	SE	t-value	p-value
a - ɛ	-121.8	34.2	-3.56	0.0082
a - i	-469.6	34.2	-13.92	<0.001*
a - ɔ	-110.7	34.2	-3.24	0.0256*
a - u	107.3	34.2	3.14	0.0358*
a - ə	38.5	34.2	1.11	0.2673
a - ɘ	19.1	34.2	0.56	0.5784
ɛ - i	-347.8	34.2	-10.26	<0.001*
ɛ - ɔ	11.1	34.2	0.32	1.0000
ɛ - u	229.1	34.2	6.68	<0.001*
ɛ - ə	160.3	34.2	4.6	0.0001*
ɛ - ɘ	140.9	34.2	4.08	0.0010*

i - ɔ	358.9	34.2	10.62	<0.001*
i - u	576.9	34.2	17.07	<0.001*
i - ə	508.1	34.2	14.78	<0.001*
i - ɘ	488.7	34.2	14.35	<0.001*
ɔ - u	218.0	34.2	6.38	<0.001*
ɔ - ə	149.2	34.2	4.29	0.0004*
ɔ - ɘ	129.8	34.2	3.77	0.0036*
u - ə	-68.8	34.2	-1.98	1.0000
u - ɘ	-88.2	34.2	-2.56	0.2218
ə - ɘ	-19.4	34.2	-0.55	1.0000

Table 2.12: Pairwise Comparison for F3 across vowel categories (a, i, u, ɛ, ɔ, ɘ and ə). The * shows the significant pairs.

Contrast	**Estimate**	**SE**	**t-value**	**p-value**
a - ɛ	-12.218	8.12	-1.505	1.0000
a - i	-11.602	8.12	-1.45	1.0000
a - ɔ	14.512	8.12	1.793	0.0733
a - u	1.135	8.12	-0.14	0.8885
a - ə	-15.014	8.12	-1.823	0.0685
a - ɘ	-14.304	8.12	-1.754	0.0797
ɛ - i	0.617	8.12	0.077	1.0000
ɛ - ɔ	26.73	8.12	3.288	0.0218*
ɛ - u	13.354	8.12	1.642	1.0000
ɛ - ə	-2.795	8.12	-0.338	1.0000
ɛ - ɘ	-2.086	8.12	-0.255	1.0000
i - ɔ	26.114	8.12	3.258	0.0242*
i - u	12.737	8.12	1.589	1.0000
i - ə	-3.412	8.12	-0.418	1.0000
i - ɘ	-2.702	8.12	-0.335	1.0000

ɔ - u	-13.377	8.12	-1.65	1.0000
ɔ - ə	-29.525	8.12	-3.581	0.0075*
ɔ - ɘ	-28.816	8.12	-3.528	0.0091*
u - ə	-16.149	8.12	-1.959	1.0000
u - ɘ	-15.439	8.12	-1.89	1.0000
ə - ɘ	0.71	8.12	0.085	1.0000

Table 2.13: Pairwise Comparison for duration across vowel categories (a, i, u, ɛ, ɔ, ɘ and ə). The * shows the significant pairs.

2.4 Conclusion

This chapter provides a comprehensive overview of Chokri's phoneme inventory. The analysis diverges from prior studies, such as those of Bielenberg and Nienu (2001). The inventory includes 34 consonants and 7 vowels. Chokri consonant makes a three-way distinction between voiced unaspirated, voiceless unaspirated, and voiceless aspirated, detailed through categories of stops, nasals, laterals, fricatives, approximants, and affricates. The vowel analysis confirms the presence of seven distinct vowels. These vowels include two front vowels [i, ɛ]), three central vowels ([a, ɘ, ə]), and two back vowels ([ɔ, u]). In the next chapter, we will discuss the phonotactic properties observed in Chokri.

References

Abramson, A. S. (1977). The VOT in the Speech of Various Languages: A Cross-Linguistic Study. In *Proceedings of the 5th International Congress of Phonetic Sciences* (pp. 164-166).

Bielenberg, B. and Nienu, Z. (2001). Chokri (Phek Dialect): Phonetics and Phonology. *Linguistics of the Tibeto-Burman Area*, 24(2).

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* [Computer program] (Version 6.4.07). Retrieved March 17, 2024, from http://www.praat.org/

Borden, G. J. and Harris, K. S. (1980). Physiology, Acoustics and Perception of Speech. *Speech science primer:* Baltimore: The Williams and Wilkins Company.

Carlson, R., Granström, B., and Fant, G. (1970). Some Studies Concerning Perception of Isolated Vowels. *Speech Transmission Laboratory Quarterly Progress and Status Report,* 1, 19–35.

Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton: The Hague.

Fant, G. (1966). A Note on Vocal Tract Size Factors and Non-Uniform F-pattern Scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report,* 7, 22–30.

Fant, G. (1975). *Speech Acoustics and Phonetics*. Springer.

Goldstein, U. (1980). *An Articulatory Model for the Vocal Tracts of Growing Children.* (PhD dissertation), Massachusetts Institute of Technology.

Gope, A. (2016). *The Phonetics and Phonology of Tone in Sylheti*. PhD dissertation. IIT Guwahati, India.

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

Hitch, D. (2017). Vowel Spaces and Systems. *Toronto Working Papers in Linguistics*. 38.

Jurafsky, D., and Martin, J. H. (2023). *Speech and Language Processing* (12th ed.). Prentice Hall.

Keating, P.A., Mikós, M.J., and Ganong, W.F. (1981). A Cross-Language Study of Range of Voice Onset Time in the Perception of Initial Stop Voicing. *Journal of the Acoustical Society of America*, *70*, 1261-1271.

Ladefoged, P. (2011). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Blackwell Pub.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages*. Blackwell Publishers, Blackwell Publishing Limited

Ladefoged, P. (1975). *A Course in Phonetics*. United States: Harcourt Brace Jovanovich.

Lodge, K. (2009). *A Critical Introduction to Phonetics.* United Kingdom: Bloomsbury Academic.

Man, C. Y. (2007*).* An Acoustical Analysis of the Vowels, Diphthongs and Triphthongs in Hakka. . *In Proceedings of the 15th International Congress of Phonetic Sciences*

Matisoff. J. A (2003) *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction.* University of California Press.

Monsen, R. B., and Engebretson, A. M. (1983). The Accuracy of Formant Frequency Measurements: A Comparison of Spectrographic Analysis and Linear Prediction. *Journal of Speech and Hearing Research*, 26, 89 - 87.

Nearey, T. M., and Assmann, P. F. (1986). Modeling the Role of Inherent Spectral Change in Vowel Identification. *The Journal of the Acoustical Society of America*, 80(5), 1297–1308.

Peterson, G. E., and Barney, H. L. (1952*).* Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184.

Pickett, J. M. (1996). *The Sounds of Speech Communication* - *A Primer of Acoustic Phonetics and Speech Perception*. Boston: Allyn and Bacon.

Rosner, B. S., and Pickering, J. B. (1994). *Vowel Perception and Production*. Oxford University Press.

Simpson, A. P. (2001). Dynamic Consequences of Differences in Male and Female Vocal Tract dimensions. *Journal of the Acoustical Society of America*, 109(5), 2153–2164.

Simpson, A. P. (2009). Phonetic Differences Between Male and Female Speech. *Language and Linguistics Compass, 3*(2), 621–640. https://doi.org/10.1111/j.1749-818x.2009.00125.x

Stevens, K. N. (1998). *Acoustic Phonetics*. MIT Press.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (1998). Perceptual Assimilation of American English Vowels by Japanese Listeners. *The Journal of the Acoustical Society of America*, 103(1), 564–574.

Teo, A. (2014). *A Phonological and Phonetic Description of Sumi, A Tibeto-Burman Language of Nagaland.* Canberra: Asia-Pacific Linguistics

Terhiija, V. and Sarmah, P. (2021). Aspiration in Voiceless Nasals in Angami. *Proceedings of meetings on acoustics Acoustical Society of America.* 42. 10.1121/2.0001403.