Chapter 4: Tone

4.1. General Introduction

In the previous chapter, we examined the phonotactic patterns and phonological processes of Chokri, focusing on how sounds are structured and modified in the language. We explored syllable constraints, noting Chokri's preference for open syllables and its adaptation of loanwords through coda deletion. Additionally, we discussed major phonological processes such as vowel deletion, lenition, and consonant alternations, all of which contribute to shaping the sound system of Chokri. This chapter will focus on exploring the tonal inventories in Chokri. We aim to identify the distinct tonal categories present in the language and analyze their acoustic properties. Additionally, we will examine how native speakers perceive and distinguish these tones.

In one of the first studies, Bielenberg and Nienu (2001) identified four tones in the Phek dialect of Chokri, viz., low, low-mid, high-mid, and high tones, along with an allotone of the high-mid tone. However, their findings were primarily based on auditory impressions, without detailed acoustic analysis. As a result, crucial aspects of tonal production and perception, such as specific acoustic features of each tone, the role of f0 height and directionality, as well as duration, remained unexplored.

Pitch variation is a universal feature in the phonetic and phonological systems of the world's languages. Although pitch is inherently an auditory phenomenon and cannot be measured directly, it correlates closely with the vibration rate of the vocal cords, i.e., the higher the frequency of vibration, the higher the perceived pitch, and vice versa. Therefore, pitch can be quantified by measuring the fundamental frequency (f0) of a sound wave (Ladefoged, 2011). Contemporary research on tonal studies emphasizes the importance of analyzing multiple acoustic parameters, particularly f0 height, directionality, slope, duration, and intensity, to understand tonal contrasts in a given language. Among these, f0 height is commonly used to quantify tonal categories, while features such as f0 directionality, slope, duration, and intensity help to distinguish tone quality, especially in the classification of level versus contour tones. In this chapter, we, therefore, aim to examine the tonal system of Chokri through a combined acoustic and perceptual approach. Specifically, we will analyze f0 height, directionality, duration, and intensity to determine the nature and number of contrastive tones in the language, and explore how native speakers perceive these tones.

4.2. Production Experiment

We conducted a production experiment with the native Chokri speakers to explore the possible tonal contrasts in this language. The following sections explain the detailed methodology adopted in this study.

4.2.1. Participants, Materials, and Recording Procedure

A dataset comprising words with five-way meaning contrasts was developed with the help of native speakers to investigate the tonal system of Chokri. This dataset included five monosyllabic sets and three disyllabic sets, each containing five homophonous words exhibiting five distinct meanings, thus forming minimal tonal sets.

Recordings were made using a Multi-Track Linear PCM Recorder (OLYMPUS LS-100) paired with a SHURE SM10 headset microphone at a sampling rate of 44,100 Hz in .wav format. Seven native speakers (2 males and 5 females) from Thipüzu village, located in Phek district, Nagaland, participated in the production experiment. Speech data were collected through three recording steps:

1. Target words spoken in isolation,
2. Target words embedded in simple priming sentences to facilitate natural tone selection by the speakers, and
3. Target words placed within a fixed sentence frame (carrier phrase): *"Say X again"* (*Vapü X si sasü te*), where *X* represents the target word.

The target words were manually randomized, and the entire dataset was repeated five times by each subject, with a considerable gap between sessions. For the final analysis, we considered only tokens from the isolation and carrier phrase contexts, resulting in a dataset of 2,800 analyzed tokens (40 words x 7 subjects x 5 repetitions x 2 contexts). Table 4.1 presents the lexical items used in the production experiment.

Chokri Word	Meaning 1	Meaning 2	Meaning 3	Meaning 4	Meaning 5
[tʃʰə]	Expression of disgust	To dig	The act of collecting dirt in a dust plate	To fetch water	Bangle

[pɘ]	Fat	To soften (rice/legumes)	Take	To hit with stone	Bridge
[sɔ]	To sow (rice)	Dry	To shake (pillar)	Count	Younger sibling
[sɘ]	To dip	Tree	To block view	Cold	Three
[ta]	To throw and splash water	Run	Chew	Walk	Mouth
[ɛlɘ]	Rest	Woodworm	Pinch	To go inward	Warmth
[mɛza]	To style	Thank-you	Friend	Catch	Weak
[mɛlɘ]	To suspect	Move	Believe	To tie a knot	To warm

Table 4.1: Dataset considered for the production experiment.

4.2.2. Measurements: Post Recording

The process of data annotation and acoustic analysis was conducted using Praat 6.1.51 (Boersma and Weenink, 2012). Each target word was extracted from the recorded utterances and saved as an individual sound file. These files were then manually segmented using a three-tier Praat TextGrid, with tiers corresponding to the word, syllable, and vowel nucleus or vowel, the latter being the carrier of pitch information.

Pitch tracks were visually examined, and pitch boundaries were manually marked at the onset and offset of each vowel. A customized Praat script (Gope, 2014, 2016) was used to extract pitch values at eleven evenly spaced time intervals at 10% increments across the total duration of each vowel segment (vowel and/or vowel+sonorous coda, if any) (from 0% onset to 100% offset).

The fundamental frequency (f0) values obtained from each token were then averaged across repetitions, including both the isolation and carrier phrase contexts. These averaged f0 trajectories were visualized and represented as line graphs using *Python* to illustrate the distinct tonal contours and facilitate comparative analysis.

4.3. Acoustic Analysis of f0 Directionality

This section presents the acoustic analysis of tonal directionality in Chokri based on pitch contours across different tonal categories. The aim is to examine how pitch movement contributes to tonal contrast in the language. By visualizing and comparing non-normalized f0 contours across speakers and tonal categories, we identify consistent tonal patterns and assess inter-speaker variations, particularly those arising from physiological differences between male and female speakers. Further, normalization techniques are employed to control for such variability and ensure accurate tonal classification. Figure 4.1 and Figure 4.2 show the speaker-wise, non-normalized pitch contours for the [ta] and [sɔ] series, respectively. These figures illustrate pitch contours as produced within the carrier frame by each speaker.

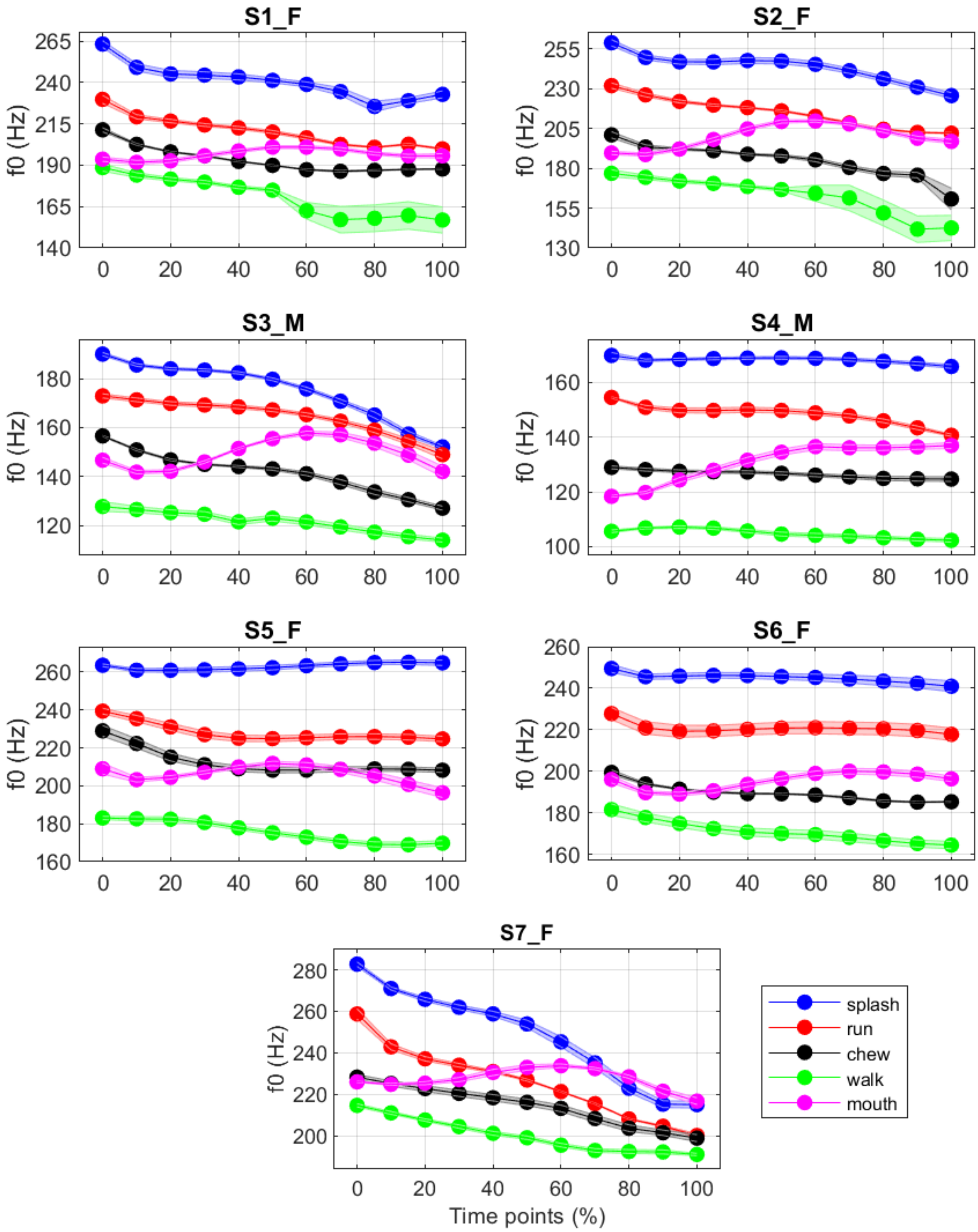


Figure 4.1: Non-normalized averaged pitch tracks for [ta] series produced by each speaker, meanings: ‘to splash water,’ ‘to run,’ ‘to chew,’ ‘to walk,’ and ‘mouth.’

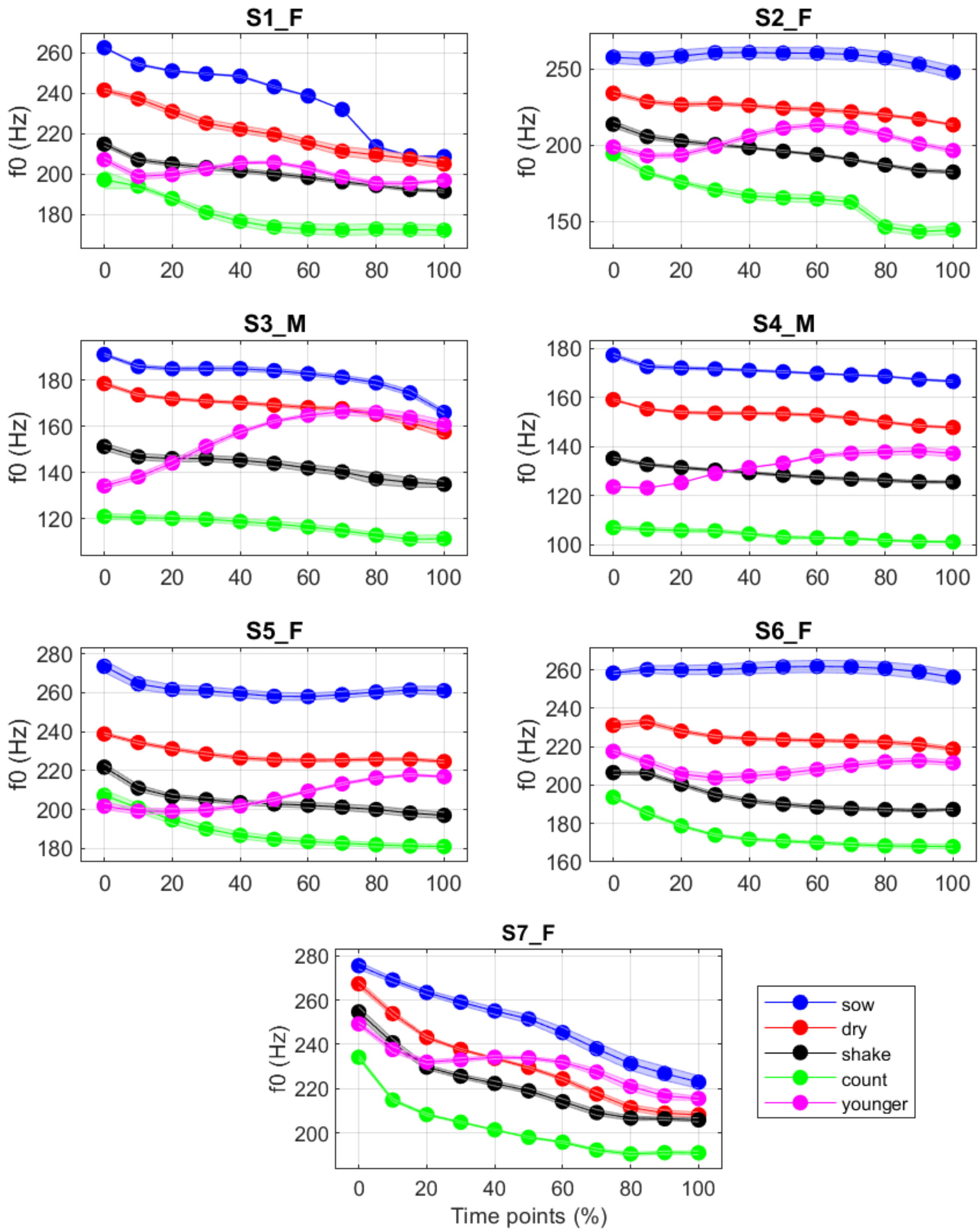


Figure 4.2: Non-normalized averaged pitch tracks for the [sɔ] series produced by each speaker, meanings: ‘to sow(paddy),’ ‘dry,’ ‘to shake/move,’ ‘to count,’ and ‘younger sibling.’

Figures 4.1 and 4.2 displayed above clearly illustrate five distinct pitch contours across two series by each subject, indicating possible five-way tonal contrasts in this language. These series further indicate that pitch varies between male and female speakers, though both groups maintain distinct tonal contrasts across tonal categories. These results suggest that the physiological differences in vocal fold structure influence the absolute pitch values but do not affect the overall tonal inventory. We also observe speaker-wise variations for each possible tone type. From the data shown above, we could preliminarily assume that there are four possible level tones and one contour tone in Chokri. At this point, we also need to decide on the tonal notations[7] that can be incorporated to explain the possible tonal contrasts in this language.

4.3.1. Chokri and its Tonal Notation

Figures 4.1 and 4.2, displayed above, indicate the presence of four level tones and one contour tone in Chokri, a tone system that seems more similar to African languages rather than the contour-rich systems of Southeast Asia. Considering the range of raw f0 values, Chokri can be effectively represented using an African-style diacritical system, and the tonal contrasts can be represented as extra high (EH), high (H), mid (M), low (L), and mid-rising (MR):

Extra High → Double acute accent (a̋)

High → Acute accent (á)

Mid → Macron (ā)

Low → Grave accent (à)

Mid-Rising → Caron (ǎ)

Figure 4.3 presents the non-normalized averaged pitch tracks of the (possible) five contrastive tones across all word series produced by all males and all females separately. This figure also gives us an opportunity to distinguish the averaged contours by speaker gender, allowing for a profound direct comparison between male and female pitch profiles and the possible tonal quality in Chokri.

[7] Tone notation is essential for documenting tonal languages, with systems varying by region. African languages, often level-tone systems (Hyman, 1979, 2011), use diacritics for high (á), mid (ā), low (à), extra-high (a̋), and extra-low (ȁ) tones, along with symbols for downstep (ꜜ) and upstep (ꜛ). Asian languages, primarily loaded with contour tones, often use Chao's (1930) 1–5 numerical system (1 = lowest, 5 = highest) to capture pitch movement, e.g., ta55 (high), ta35 (high rising), ta214 (complex) (Yip, 2002). American tonal languages, on the other hand, prefer to use numbers, albeit in a reverse order,where, 5 = lowest, 1 = highest, with single digits for level tones and hyphenated pairs for contours (e.g., si1 high, si3-2 high rising).

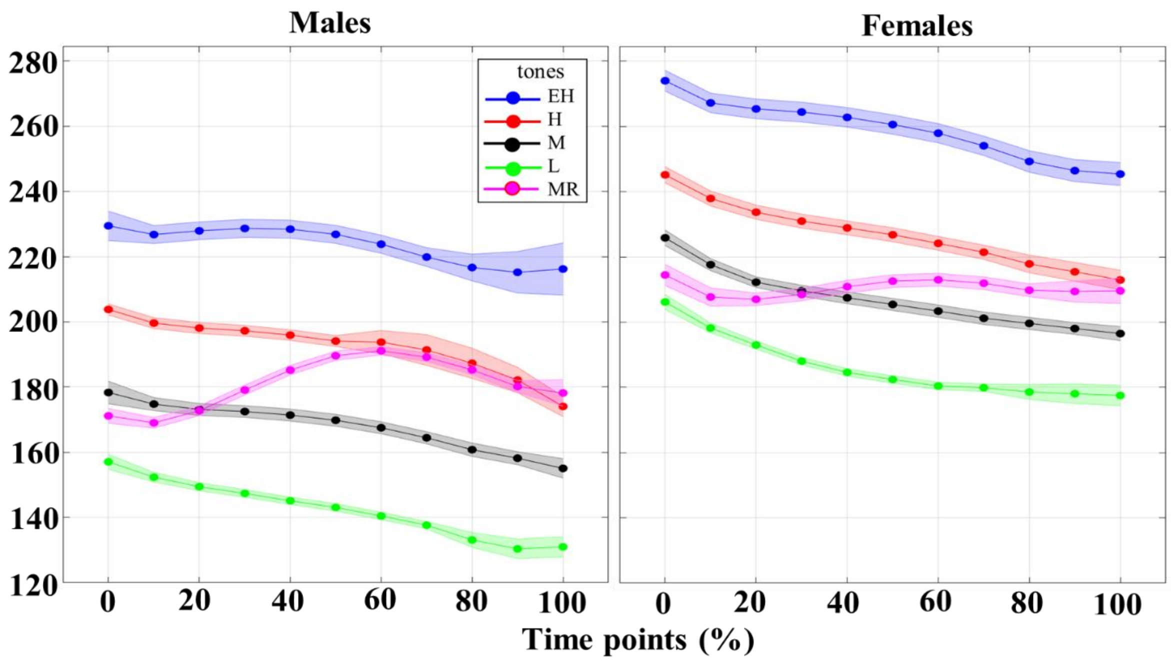


Figure 4.3: Averaged non-normalized pitch tracks for all the words produced by all males (left panel) and all females (right panel) separately.

These findings confirm that while male and female speakers differ in absolute f0 due to physiological factors, they both produce the five contrastive tones with consistent and distinguishable pitch patterns. The presence of these five distinct tonal contours across genders strongly supports a possible five-way tonal contrast in Chokri.

4.3.2. Normalization of Speech Data

Biological differences, particularly in vocal fold structure and vocal tract dimensions, have a substantial influence on pitch production. As Simpson (2009) notes, male vocal folds are typically longer and thicker than those of females, resulting in slower vibration and, consequently, lower pitch. Female vocal folds, being shorter and thinner, vibrate more rapidly and yield higher pitch values. Ladefoged (2011) similarly explains that pitch is directly related to the rate of vocal fold vibration, i.e., slower rates produce lower pitches, and faster rates produce higher ones.

These anatomical differences pose challenges for cross-speaker comparisons in acoustic studies, especially when attempting to identify phonologically relevant patterns that transcend individual physiological variability. The analysis of non-normalized pitch tracks (Figure 4.3) reveals clear differences in pitch range between male and female speakers. Male speakers consistently exhibit a lower f0 across tonal categories, whereas female speakers demonstrate elevated pitch profiles. This disparity, though expected, necessitates normalization to avoid misinterpretation of the tonal inventory. For a clearer understanding, the average f0 spacing of the four level tones, comparing tonal intervals across genders, is shown in Table 4.2.

Tone	Male	Female
Extra High vs. High	30 Hz	34 Hz
High vs. Mid	26 Hz	22 Hz
Mid vs. Low	27 Hz	24 Hz

Table 4.2: Average f0 spacing between level tones in male and female speech.

It is also observed that, on average, female speakers' f0 values are approximately 45 Hz higher than those of male speakers. Without normalization, such variability could distort acoustic patterns and obscure phonological generalizations. To mitigate these effects and ensure a more accurate representation of Chokri's tonal system, a *z-score normalization* procedure was implemented using the formula:

$$z = \frac{(\boldsymbol{x} - \boldsymbol{\mu})}{\sigma}$$

where, x is the raw pitch value, μ is the speaker-specific mean, and σ is the standard deviation. This method standardizes the data by accounting for inter-speaker and intra-speaker pitch variations, enabling valid comparison across speakers/genders and preserving the integrity of tonal distinctions in the language.

Figures 4.4 and 4.5 below present the normalized speaker-wise pitch tracks for the [sɔ] and [ta] series, respectively. The visual observations clearly hint at the presence of five contrastive pitch contours across speakers. The patterns are consistent and replicable, affirming the robustness of the five-way tonal contrasts in Chokri.

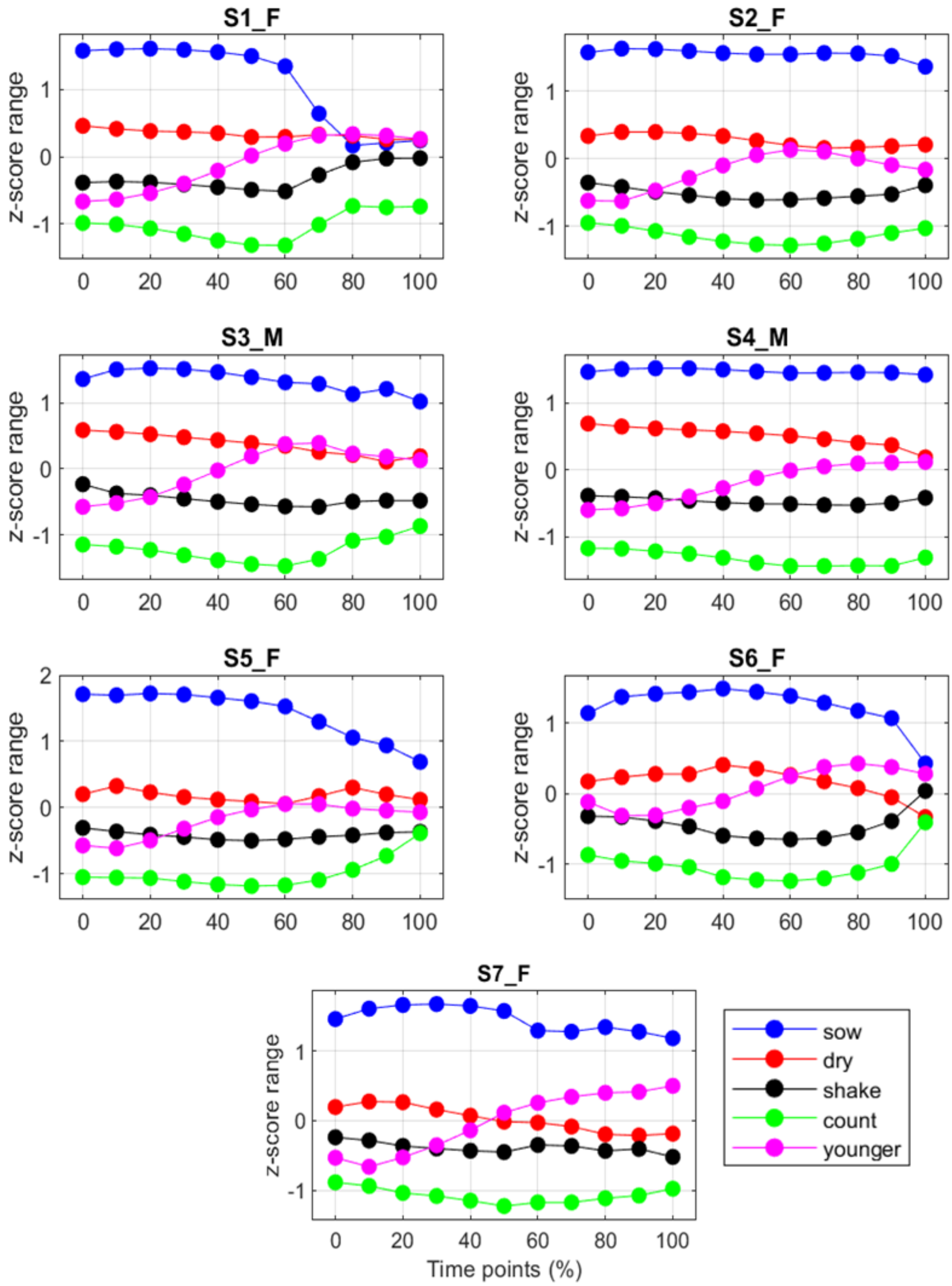


Figure 4.4: Normalized and averaged (across five repetitions by each speaker) pitch tracks for [sɔ] series produced by individual speakers.

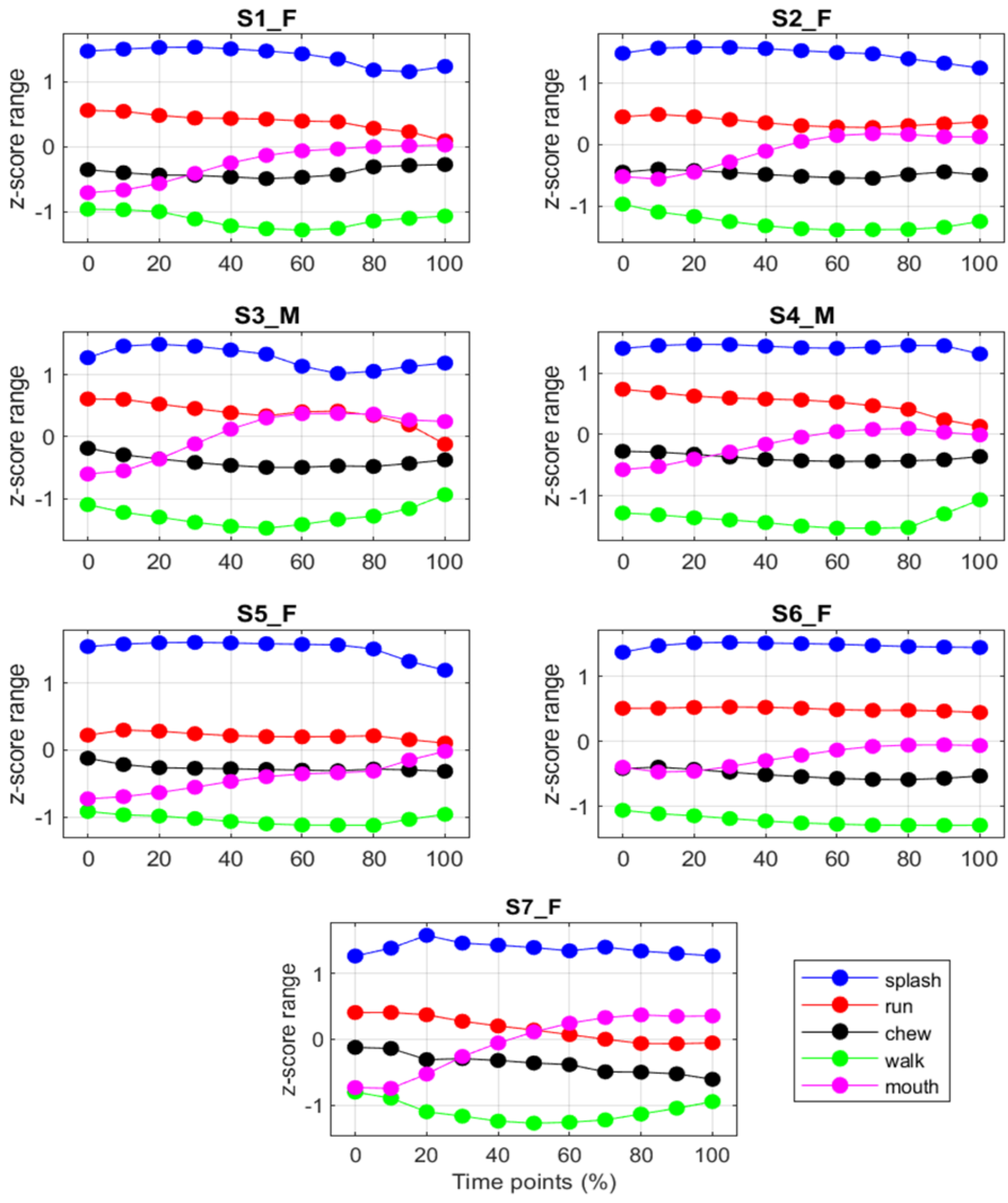


Figure 4.5: Normalized and averaged (across five repetitions by each speaker) pitch tracks for [ta] series produced by individual speakers.

4.3.3. Statistical Modeling

The visual observations of the Figures 4.1 – 4.5 indicate towards a five-way tonal contrast in Chokri. However, to confirm the visual observations and to determine the quantity and quality of contrastive tones in Chokri require a regious statistical analysis. Usually, traditional statistical approaches, such as analysis of variance (ANOVA), linear models (LMs), and generalized linear models (GLMs), are commonly used to determine the possible tonal contrasts in a given language. ANOVA and LMs are typically applied when the predictors are categorical, and the outcome variables are continuous, providing a straightforward method for testing mean differences across groups. GLMs extend these approaches by accommodating non-normal distributions of the response variable and using link functions, making them applicable to a wider variety of data types. Mixed-effects models have become increasingly popular, particularly when accounting for speaker variability in repeated-measures designs, with the *lme4* package in R frequently used for mixed-effects modeling.

Additionally, non-parametric tests, such as the Kruskal–Wallis test, are used in prosodic research when the assumptions of normality or homogeneity of variance for parametric tests are violated. The Wilcoxon rank-sum test (Mann–Whitney U test) is also employed for comparisons between two groups, particularly when their distributions are skewed. For repeated measures designs, Friedman tests are used to analyze within-speaker variations in prosodic parameters across conditions (Sonderegger, 2023).

Despite their widespread use, these traditional models have limitations when applied to acoustic data. Linear models and ANOVA assume linearity and homogeneity of variance, assumptions often violated in real-world acoustic datasets. These methods also struggle to capture the smooth, non-linear transitions that characterize many prosodic phenomena, such as the gradual rise or fall in pitch or intensity over a syllable or word. While GLMs offer more flexibility, they still require the user to specify the functional form of each predictor, which may not be known a priori or may oversimplify the actual relationship. Although mixed-effects models can handle nested structures, they are typically limited to linear relationships, which may not be appropriate for data exhibiting complex, non-linear trends.

Generalized Additive Models (GAMs) is one of the most prefered standard tool in experimental phonetics for modeling continuous, non-linear data such as pitch trajectories. By using smooth functions, GAMs can capture the complex temporal patterns of prosodic features while flexibly accommodating multiple interacting variables (e.g., tone category, syllable position, vowel type) without imposing restrictive parametric assumptions. Their ability to model non-linear relationships, handle unbalanced datasets, and incorporate random effects makes them well suited for the analysis of tonal data. The growing body of work employing GAMs, particularly in prosody, reflects their status as a current best practice for analyzing pitch and other time-varying acoustic measures.

The term "smooth functions" refers to a flexible, data-driven approach for modeling the relationship between predictors and response variables. In contrast to traditional linear models, which assume a strictly linear or stepwise relationship, the GAM model analyzes relationships more flexibly. For example, in a traditional linear model, the relationship between f0 and tone might be written as $f0 = a + b \times tone$, where "a" represents the intercept (the baseline value of f0 when the predictor variable, tone, is at its reference level), and "b" represents the slope (the change in f0 associated with each unit change in the predictor). This assumes that each tone category shifts the f0 value by a fixed amount relative to a baseline. However, this approach may oversimplify the relationship, as acoustic data often exhibit more complex, non-linear patterns. In contrast, a GAM models the same relationship as $f0 = s(tone)$, where s(tone) represents a smooth curve estimated from the data rather than a fixed coefficient for each tone. This flexibility is especially valuable in phonetics, where acoustic parameters typically exhibit non-linear and gradient variations. If we want to model how f0 evolves over time, for instance, we could specify $f0 = s(time)$ or $f0 = s(time, by = tone)$, where the latter allows the shape of the f0 contour to vary by tonal category, capturing the dynamic trajectory of the mid-rising tone across time.

In the present study, GAMs were selected as the primary statistical tool to analyze the relationship between tonal contrasts and three acoustic parameters. The decision to use GAMs was based on preliminary visualizations of f0 contours sampled at 10% intervals across normalized time, which revealed that tonal patterns, particularly for mid-rising tone, followed continuous, curved contours rather than stepwise or linear shifts. These findings, characteristic of tonal systems in many languages, suggested that a linear model would be insufficient to capture the underlying dynamics. Consequently, GAMs were chosen for their ability to model complex, potentially non-linear relationships through smooth functions.

We employed a simplified GAM model specification in this study: f0 ~ tone, intensity ~ tone, and duration ~ tone, treating tone as a categorical predictor in each case. These models were fitted to the mean raw (non-normalized) values of each acoustic parameter for two main reasons: first, while f0 varies dynamically over time, intensity and duration are scalar measures that do not exhibit temporal variation; and second, using mean f0 values allowed for a consistent, comparative analysis across tones while maintaining uniformity across all three acoustic parameters.

The data were processed using R, and possible tonal categories were mapped as EH, H, M, L, and MR for our understanding at this point. GAM models were computed using the mgcv package, with a Gaussian family and 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.

4.4. Results:

4.4.1. f0 Contrasts in Monosyllabic Words

The GAM model for f0 analysis demonstrated significant variability across tone categories ($p < 0.001$), with an adjusted R-squared value of 0.331, explaining 33.4% of the variance. The estimated marginal means (EMMs) for each tone category are presented in Table 4.3.

Tone	**Estimated f0 (Hz)**	**95% CI (Lower)**	**95% CI (Upper)**
EH	237	232	242
H	205	200	210
M	186	180	191
L	163	157	168
MR	193	188	198

Table 4.3 Estimated Marginal Means (EMMs) and 95% Confidence Intervals (CI) for fundamental frequency (f0) across tone categories (EH, H, M, L, and MR).

The analysis revealed significant differences in f0 across all tonal categories, with EH exhibiting the highest f0 and L the lowest. The greatest contrast was observed between EH and L ($t = 20.097$, $p < 0.001$), confirming a distinct pitch separation between these two tones. To explore which specific tone contrasts contributed to this variability, a pairwise comparison was conducted, and the results are summarized in Table 4.4.

Contrast	**f0 (Hz)**	**t-value**	**p-value**
EH-H	32.27	8.72	<0.0001*
EH-M	51.43	13.90	<0.0001*

EH-L	74.33	20.09	<0.0001*
EH-MR	44.30	11.97	<0.0001*
H-M	19.16	5.18	<0.0001*
H-L	42.06	11.37	<0.0001*
H-MR	12.02	3.25	<0.0002*
M-L	22.90	6.19	<0.0001*
M-MR	-7.14	-1.92	<0.0014*
L-MR	-30.04	-8.12	<0.0001*

Table 4.4: Pairwise comparisons for estimated f0 across tonal contrasts (EH, H, M, L, and MR). The asterisk (*) indicates the significant pairs.

The pairwise comparisons reported in Table 4.4 underscore a clear pattern for estimated f0 across the five contrastive tones and reveal a consistent pattern of significant differences among all the tonal pairs. All comparisons involving the EH tone show statistically significant differences with the other four types. Specifically, EH exhibits a consistently higher f0 across all other tones, with the largest differences observed relative to L (74.33 Hz), followed by M (51.43 Hz), MR (44.30 Hz), and H (32.27 Hz). The statistical significance ($p < 0.0001$) across all comparisons involving EH confirms its prominence in terms of pitch height.

Similarly, the H tone is significantly higher than M, L, and MR, with differences of 19.16 Hz, 42.06 Hz, and 12.02 Hz, respectively. These differences are also observed to be statistically significant ($p < 0.0002$). The M tone also exhibits significant differences from both L and MR, being 22.90 Hz higher than L and 7.14 Hz higher than MR, with both comparisons reaching statistical significance ($p < 0.0001$ and $p = 0.0014$, respectively). The comparison between L and MR further illustrates the pitch hierarchy, with MR significantly higher than L by 30.04 Hz ($p < 0.0001$).

The findings, indeed, confirm our visual observations that there are five contrastive tones in Chokri in which f0 directionality serves as a robust cue for distinguishing tonal quality in this language. The f0 contours (Figures 4.1-4.5) therefore, confirms that there are four level tones, viz., extra high, high, mid, and low and a contour tone, viz., mid-rising in Chokri. This analysis reveals a well-structured tonal hierarchy, with the following decreasing order of f0: EH > H >

M > MR > L. A visual representation of the f0 distribution across the five tonal categories in monosyllabic words is provided in Figure 4.6.

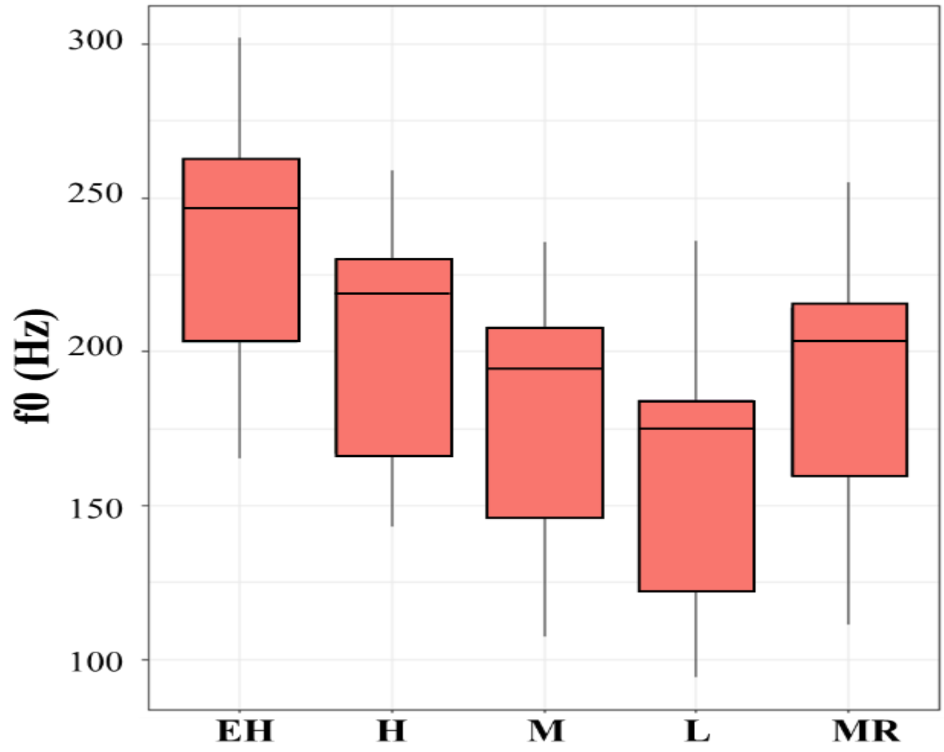


Figure 4.6: The range of mean f0 (f0 height), averaged across all the tokens produced by all the subjects for the five contrastive tones observed in monosyllabic words.

4.4.2. The Role of Duration in Monosyllabic Words

The estimated marginal means for duration are presented in Table 4.5. Notably, the results indicate that the MR tone has a significantly longer duration compared to the other tones, especially in contrast to the L tone ($t = -10.923$, $p < 0.001$). However, no significant differences were observed in the duration between most other tone pairs, suggesting that duration may be a weaker differentiator among the tones relative to f0. The pairwise comparison for duration across the five tones is summarized in Table 4.5.

Tone	Estimated Duration (ms)	95% CI (Lower)	95% CI (Upper)
EH	153	146	160
H	157	150	164
M	159	152	166
L	143	137	150
MR	198	191	204

Table 4.5: Estimated Marginal Means (EMMs) and 95% Confidence Intervals (CI) for duration across tone categories (EH, H, M, L, and MR).

Contrast	Duration (ms)	t-value	p-value
EH-H	-3.46	-0.69	1.000
EH-M	-5.40	-1.08	1.000
EH-L	9.92	2.00	0.4559
EH-MR	-44.21	-8.92	<0.0001*
H-M	-1.94	-0.39	1.000
H-L	13.38	2.70	0.707
H-MR	-40.75	-8.22	<0.0001*
M-L	15.32	3.09	0.206
M-MR	-38.82	-7.83	<0.0001*
L-MR	-54.14	-10.92	<0.0001*

Table 4.6: Pairwise Comparisons for estimated duration across tonal contrasts (EH, H, M, L, and MR). The asterisk (*) indicates the significant pairs.

The pairwise comparisons displayed in Table 4.6 expectedly reveal significant differences primarily involving the MR tone, which is consistently longer than the other four level tones. Specifically, the contrast between EH and MR shows a significant duration difference of 44.21 milliseconds ($p < 0.0001$), with MR being notably longer. MR also differs significantly from the H, M, and L tones, showing longer durations by 40.75 ms, 38.82 ms, and 54.14 ms, respectively, with all differences reaching statistical significance ($p < 0.0001$). These results suggest that the MR tone is distinctively longer in duration across the tonal contrasts. Figure 4.7 shows the mean duration (measured in milli seconds: ms) of all the contrastive tones averaged across all the tokens produced by all the subjects.

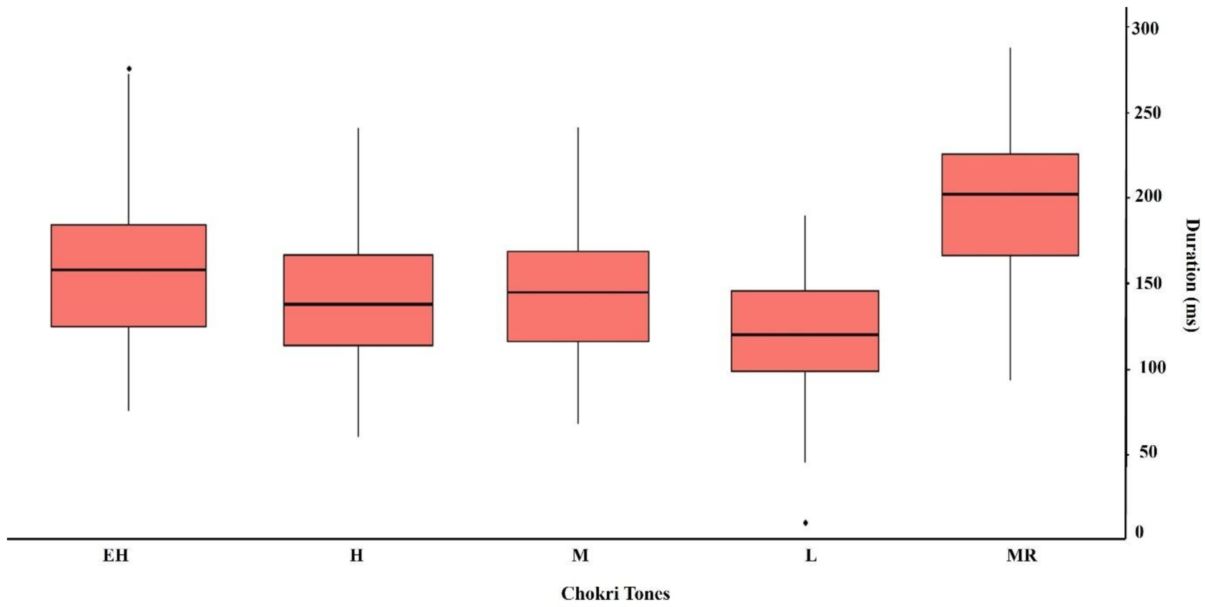


Figure 4.7: Mean duration (measured in ms), averaged across all the tokens produced by all the subjects for the five contrastive tones observed in monosyllabic words.

As seen in the above figure, the Mid Rising tone stands out as the most significant in terms of duration compared to the other tones. With a median duration of around 200 ms, it is noticeably longer than the four level tones: EH, H, M, and L, which all fall between approximately 125 ms and 150 ms. Its wider interquartile range, from about 160 ms to 230 ms, shows greater variability, reflecting the dynamic nature of this tone. Unlike the level tones that maintain a steady pitch and thus require less time, the Mid Rising tone involves a pitch movement from mid to high, which naturally demands a longer articulatory process. This extended duration highlights the complexity of the mid-rising tone, as it requires more time to complete the tonal transition, making it distinctly longer and more variable than the more stable and shorter level tones.

4.4.3. The Role of Intensity in Monosyllabic Words

The GAM model for intensity yielded an adjusted R-squared value of 0.234, explaining 23.7% of the variance. While overall differences in intensity were observed across the tones, no significant pairwise contrasts were found. The estimated intensity values for each tone are very similar, indicating minimal variation in intensity. The estimated marginal means for intensity are provided in Table 4.7, and the pairwise comparisons are summarized in Table 4.8.

Tone	Estimated Intensity (dB)	95% CI (Lower)	95% CI (Upper)
EH	68.3	67.7	68.9
H	67.9	67.3	68.6
M	66.6	66	67.3
L	65	65.4	66.7
MR	67.9	67.3	68.5

Table 4.7: Estimated Marginal Means (EMMs) and 95% Confidence Intervals (CI) for Intensity across tone categories (EH, H, M, L, and MR).

Contrast	Intensity (dB)	t-value	p-value
EH-H	0.378	0.83	1.000
EH-M	1.673	3.71	0.21
EH-L	6.288	13.97	1.000
EH-MR	0.403	0.89	1.000
H-M	1.296	2.87	0.408
H-L	5.910	13.13	1.000
H-MR	0.025	0.05	1.000
M-L	4.615	10.25	1.000
M-MR	-1.271	-2.82	0.485
L-MR	-5.886	-13.08	1.000

Table 4.8: Pairwise Comparisons for estimated intensity across tonal contrasts (EH, H, M, L, and MR). No pairs were observed to be significant in terms of intensity values.

The pairwise comparisons displayed in Table 4.8 for estimated intensity across the five tones show that, overall, there are no statistically significant differences in intensity between any of the tone pairs. While some contrasts exhibit relatively large differences in mean intensity for instance, EH is 6.29 dB higher than L, and H is 5.91 dB higher than L none of these differences reach statistical significance after adjustment for multiple comparisons. Among the contrasts, the comparison between EH and M shows a difference of 1.67 dB with a t-value of 3.71, and the M – MR comparison shows a difference of -1.27 dB with a t-value of -2.82. Though these may appear noteworthy in raw values, the associated p-values (0.21 and 0.485, respectively) indicate that these differences are not statistically significant. The lack of significant pairwise differences suggests that intensity is not a robust cue for distinguishing among these tonal contrasts in this current study. Although the Low tone tends to have a lower mean intensity compared to other tones, the variability is high enough that these differences do not reach the threshold for statistical significance. Thus, unlike f0 or duration, intensity does not appear to systematically differentiate the tonal contrasts examined in this study. Figure 4.8 presents the mean intensity (calculated in dB), averaged across all the tokens produced by all the speakers for each tone type.

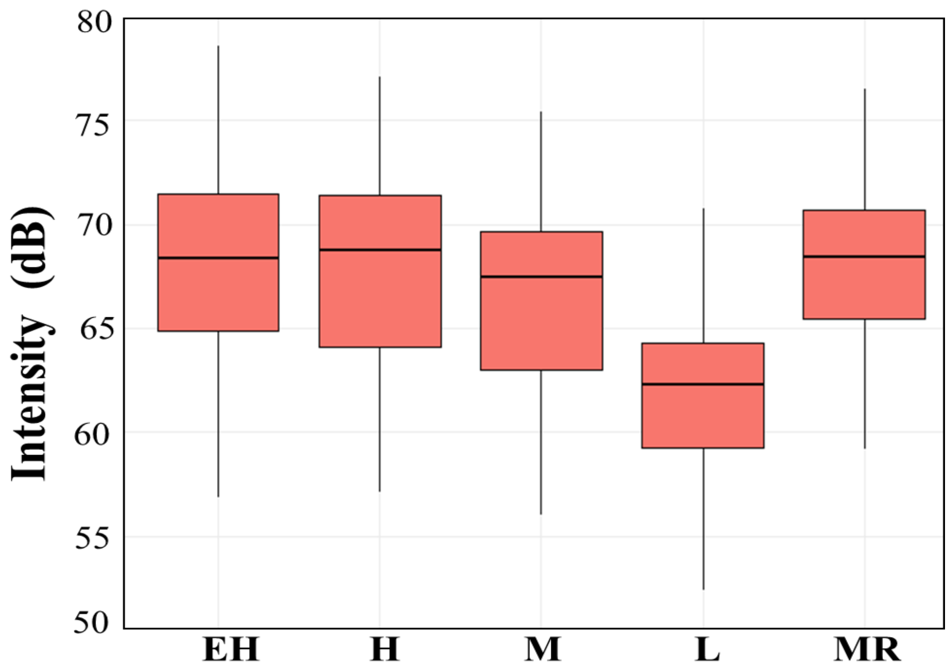


Figure 4.8: Mean intensity (measured in dB), averaged across all the tokens produced by all the subjects for the five contrastive tones observed in monosyllabic words.

A closer examination of tonal distribution across lexical categories reveals notable asymmetries between nominal and verbal roots. In monosyllabic nouns, all five tonal categories, viz., extra high, high, mid, low, and mid-rising, are attested. This unrestricted tonal occurrence suggests that tone is a lexically contrastive feature in nominal roots and plays a vital role in distinguishing lexical meaning. Such richness of tonal expression in nouns aligns with typological observations in other tone languages, where nominal categories tend to exhibit a wider tonal range than verbal forms (Yip, 2002; Hyman, 2007).

Conversely, in monosyllabic verbs, tonal distribution appears a bit constrained. Only the four level tones (EH, H, M, L) are observed, while the contour Mid-Rising tone is notably absent. This restriction supports a cross-linguistic tendency where contour tones are generally disfavoured in shorter prosodic domains unless supported by specific phonological environments (Michaud, 2004). The absence of contour tones in monosyllabic verbs implies structural constraints on the tone-bearing units (TBUs) in verbal morphology, likely due to their shorter duration or prosodic weight.

Tables 4.9 and 4.10 show the tonal distribution in monosyllabic nominal and verbal roots, respectively.

Tone	Monosyllabic Nominal roots	Gloss
Extra High (EH)	[pɛ̋]	Mushroom
High (H)	[dʒə́]	Water
Mid (M)	[zɛ̄]	Machete
Low (L)	[fə̀]	Shawl
Mid Rising (MR)	[pə᷄]	Bridge

Table 4.9: Tonal distribution in monosyllabic nominal roots.

Tone	Monosyllabic	Gloss
Extra High (EH)	[ŋɔ̋]	See
High (H)	[pʰú]	Search
Mid (M)	[tʃə̄]	Sprout
Low (L)	[tʃʰə̀]	Fetch
Mid Rising (MR)	-	-

Table 4.10: Tonal distribution in monosyllabic verbal roots.

These results suggest that while the tonal inventory of Chokri is phonologically rich, it interacts differently with grammatical categories. The unrestricted use of contour tones in nouns contrasts with their absence in verbs, hinting at a deeper morphophonological conditioning of tone. This distinction will be further explored in the subsequent sections on disyllabic words and in Chapter 6.

4.4.4. Tonal Contrasts in Bisyllabic Words

The previous sections established that Chokri exhibits a five-way tonal contrast in monosyllabic words, with notable differences in tonal distribution between nouns and verbs. Extending this analysis, the current section explores the tonal patterns in bisyllabic roots. This progression is critical, as bisyllabic forms offer a broader prosodic domain in which tonal phenomena can manifest, including the possible realization of contour tones in previously restricted environments.

4.4.4.1. f0 Contrasts in Bisyllabis Words

To determine the tonal patterns in bisyllabic roots, a methodology parallel to that used for monosyllables was adopted. This detailed temporal sampling enabled a fine-grained analysis of tonal trajectories across syllables. The resulting pitch values were normalized across speakers and plotted to observe tonal realizations (Figures 4.9-4.12).

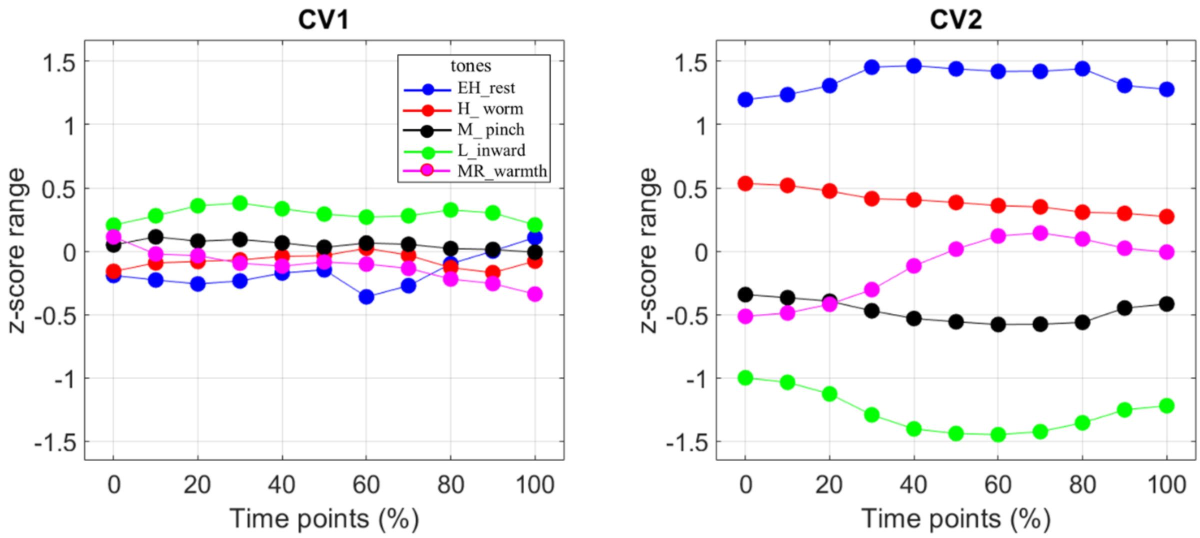


Figure 4.9: Normalized pitch-contours, averaged across all the tokens produced by all the speakers) of [ɛlɘ] series, meaning ‘to rest [EH],’ ‘wood worm H],’ ‘pinch [M],’ ‘inward [L],’ and ‘warmth [MR].’

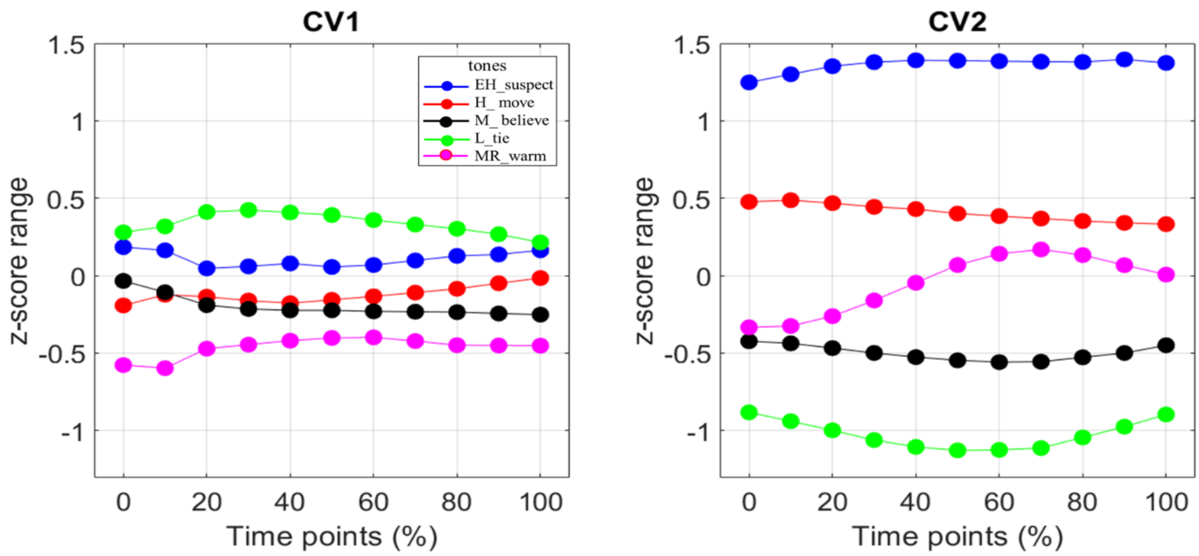


Figure 4.10: Normalized pitch-contours, averaged across all the tokens produced by all the speakers) of [mɛlɘ] series, meaning ‘to suspect [EH],’ ‘to move [H],’ ‘believe [M],’ ‘to tie [L],’ and ‘to heat/warm [MR].’

These visual observations of the Figures 4.9 – 4.10 (above) and 4.11 – 4.12 (below) revealed that, as with monosyllables, bisyllabic roots in Chokri exhibit a five-way tonal contrast. However, the contrastive tones are exclusively realized on the final syllable, while the initial syllable is uniformly assigned a mid tone. This pattern is consistently observed across all the words considered in this study (see Figs. 4.9 – 4.12).

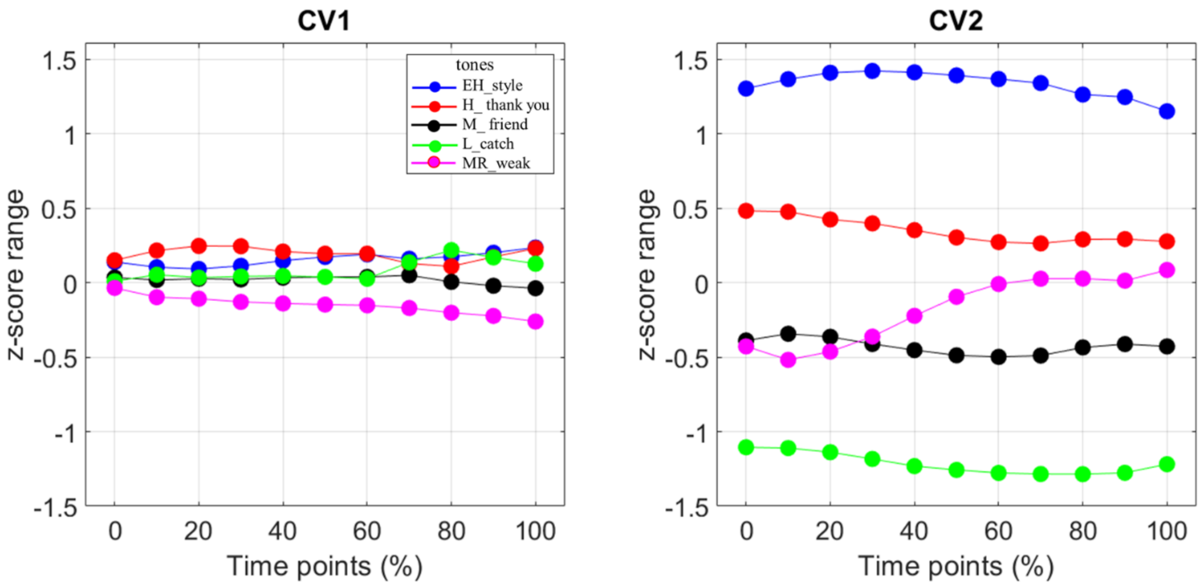


Figure 4.11: Normalized pitch-contours, averaged across all the tokens produced by all the speakers) of [mɛza] series, meaning ‘to style [EH],’ ‘thank you [H],’ ‘friend [M],’ ‘catch [L],’ and ‘weak [MR].’

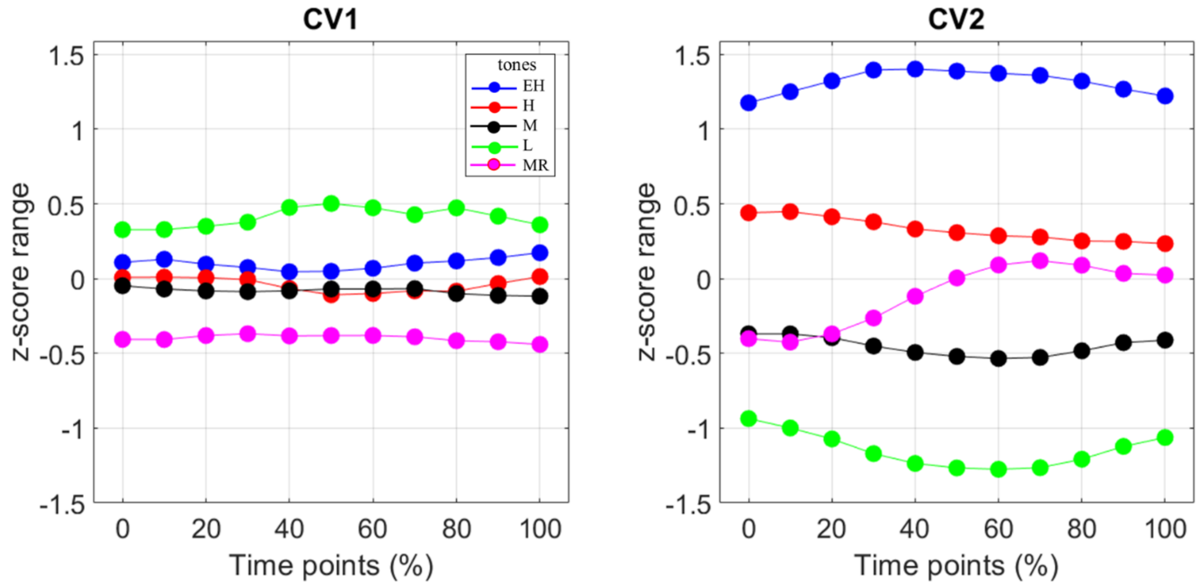


Figure 4.12: Normalized pitch-contours, averaged across all the tokens across all the contrastive tones, produced by all the speakers

Figure 4.12 summarizes the averaged normalized pitch tracks for all bisyllabic words produced by all the speakers across all the tones, reinforcing the consistent pattern of mid tone on the first syllable and contrastive tone on the second.

This asymmetric distribution reveals a hierarchy of tonal prominence, whereby the *final syllable* serves as the primary site for tonal contrast in this language. In contrast, the initial syllable functions as a tonally neutral position, consistently realized with a mid tone. The Generalized Additive Model (GAM) for f0 at the CV1 position showed no significant differences between the contrastive tonal pairs, as all p-values exceeded the 0.05 threshold. The model's adjusted R-squared value was negative (-0.00147), suggesting that it explains very little variation in the data. The estimated marginal means for f0 at CV1 are shown in Table 4.11, which reveal minimal variation across tonal contrasts: EH and H both average 187 Hz, M is slightly lower at 186 Hz, L is marginally higher at 190 Hz, and MR registers at 185 Hz. The narrow 95% confidence intervals across all estimates and the overlap between tones further confirm the lack of statistically meaningful differences. As the statistical model and confidence intervals both indicate little to no tonal separation at CV1, no further pairwise comparisons were conducted.

Tone	Estimated f0 (Hz)	SE	95% CI (Lower)	95% CI (Upper)
EH	187	2.45	182	191
H	187	2.38	183	192
M	186	2.44	181	191
L	190	2.38	185	195
MR	185	2.38	180	190

Table 4.11: Estimated Mean values of f0 (Hz) at CV1 for the five tones (EH, H, M, L, and MR), along with their standard errors (SE) and 95% confidence intervals (CI).

In contrast, the GAM model for f0 in CV 2 reveals clear and statistically significant distinctions among the five contrastive tones. The model showed that f0 varied systematically with tone, with the EH tone consistently exhibiting the highest f0 values. The estimated differences between EH and the other tones are significant: 29.24 Hz higher than H, 49.42 Hz higher than M, 69.41 Hz higher than L, and 37.41 Hz higher than MR, all with high statistical significance ($p < 0.0001$). The adjusted R-squared value of 0.308, with 31% deviance explained, suggests that the model captures a meaningful portion of the variation in f0 at this syllable position. Pairwise comparisons in Table 4.14 further support these findings, showing significant differences between all tone pairs. Notably, H is significantly higher than M (20.18 Hz), L (40.17 Hz), and MR (8.18 Hz), all with p-values below 0.05. M also differs significantly from both L (19.99 Hz) and MR (-12 Hz), while the contrast between L and MR is the largest after EH, with L being 32 Hz higher than MR. These results indicate a robust hierarchical structure in pitch across the tonal categories at CV2, with each tone occupying a distinct position in the f0 space.

Contrast	f0 (Hz)	SE	t-value	p-value
EH-H	29.24	3.37	8.67	<.0001*
EH-M	49.42	3.41	14.47	<.0001*
EH-L	69.41	3.37	20.57	<.0001*
EH-MR	37.41	3.37	11.09	<.0001*
H-M	20.18	3.37	5.99	<.0001*
H-L	40.17	3.32	12.08	<.0001*
H-MR	8.18	3.32	2.45	0.0148*
M-L	19.99	3.37	5.93	<.0001*

M-MR	-12	3.37	-3.56	0.0038*
L-MR	-32	3.33	-9.61	<.0001*

Table 4.12: Pairwise Comparisons for estimated f0 across tonal contrasts (EH, H, M, L, and MR) at CV2. The asterisk (*) shows the significant pairs.

To complement the statistical analysis, Figure 4.13 presents a box plot that visually illustrates the f0 height distribution across the five tonal categories for both CV1 and CV2 positions in bisyllabic words. This visual representation highlights the systematic differences in f0 among the contrastive tones, clearly showing the elevated f0 values for EH and the distinct separation between tonal categories at each syllable position. The box plot summarizes the tonal hierarchy, reinforcing the patterns identified in the GAM model and pairwise comparisons.

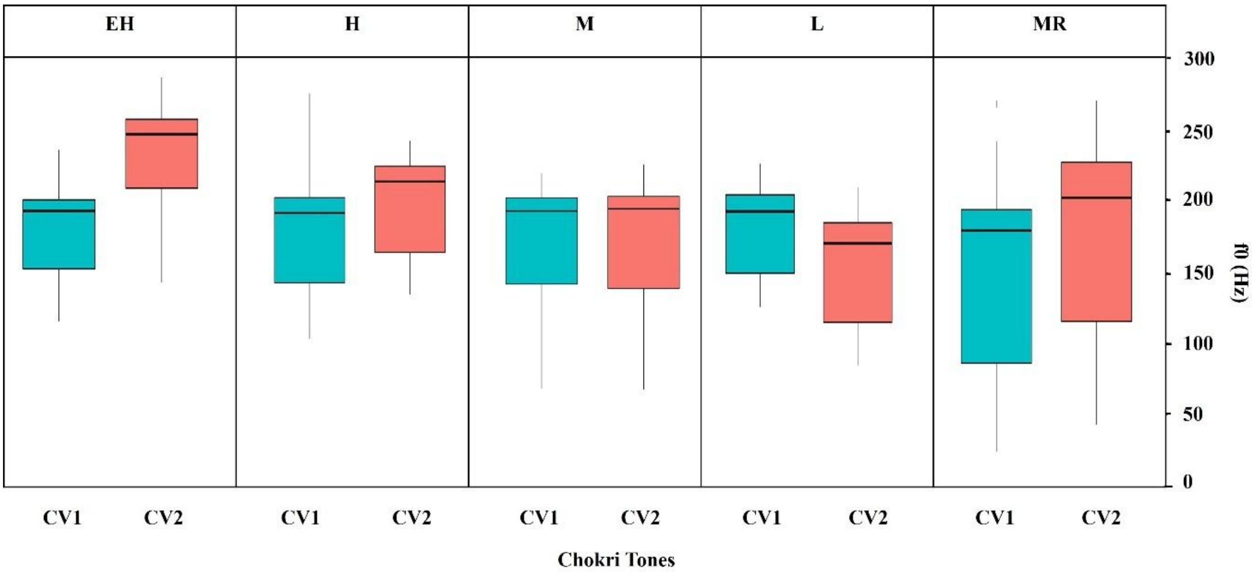


Figure 4.13: Mean f0 (f0 height), averaged across all the tokens produced by all the subjects for the five contrastive tones observed in bisyllabic words.

4.4.4.2. The Role of Duration in Bisyllabic Words

This hierarchical structure of tone in Chokri is further corroborated by duration measurements (see Figure 4.14). The relationship between tone and duration has been widely discussed in the literature on tonal phonology. For example, Zhang (2002) argues that low or non-prominent tones often exhibit reduced duration, while contrastive or contour tones tend to be phonetically longer. Thus, the duration asymmetry between initial and final syllables in Chokri provides strong empirical support for a rule-governed tonal grammar, where the final syllable serves as the main site for tonal contrast, while the initial syllable is assigned a mid tone as a neutral default.

Similar to the f0 analysis shown in Table 4.12, the estimated mean duration values at the CV1 position across the five tonal categories, as determined by the Generalized Additive Model (GAM), exhibit only modest variation. The confidence intervals overlap substantially, indicating no reliable differences between the tones (see Table 4.13). At CV1, the EH and H tones have the shortest durations at 95.1 ms and 93.7 ms, respectively. The M and L tones have slightly longer estimated durations of 99.6 ms and 101.8 ms, while the MR tone falls between these extremes at 96.5 ms. Despite these numerical differences, the small standard errors and overlapping 95% confidence intervals for all tones suggest that there is no statistically significant separation between them. This pattern implies that duration is not a consistent acoustic cue for distinguishing between tonal categories at CV1. As a result, pairwise comparisons between tonal contrasts were not conducted.

Tone	**Estimated Duration (ms)**	**SE**	**95% CI (Lower)**	**95% CI (Upper)**
EH	95.1	2.39	90.4	99.8
H	93.7	2.32	89.2	98.3
M	99.6	2.39	94.9	104.3
L	101.8	2.33	97.3	106.4
MR	96.5	2.33	91.9	101

Table 4.13: Estimated mean values of duration (ms) at CV1 for the five tones (EH, H, M, L, and MR), along with their standard errors (SE) and 95% confidence intervals (CI).

While there were no significant differences in duration at the CV1 position, the duration at CV2 reveals a clear pattern of tonal variation, particularly with the MR tone. Pairwise comparisons indicate that the MR tone is significantly longer than all other tones, with differences ranging from -45.25 ms to -78.10 ms, all of which are highly significant (p-values < 0.0001). This finding highlights the MR tone as distinctively longer in duration at this syllable position compared to the level tones. Although some numerical contrasts among the level tones (EH, H, M, and L) appear substantial—such as the 32.85 ms difference between H and L, none of these differences reach statistical significance after correction. This suggests that while the L tone tends to be longer, duration does not reliably differentiate the level tones. Overall, the duration at CV2 serves as a robust acoustic marker for the MR tone but does not statistically differentiate among the four level tones. The pairwise differences are detailed in Table 4.14.

Contrast	Duration (ms)	SE	t-value	p-value
EH-H	-9.51	5.21	-1.82	0.68
EH-M	-2.14	5.28	-0.40	1
EH-L	23.34	5.21	4.47	0.1
EH-MR	-54.76	5.21	-10.50	<.0001*
H-M	7.37	5.2	1.41	1
H-L	32.85	5.14	6.39	0.1
H-MR	-45.25	5.14	-8.80	<.0001*
M-L	25.48	5.21	4.89	0.1
M-MR	-52.62	5.21	-10.10	<.0001*
L-MR	-78.1	5.14	-15.18	<.0001*

Table 4.14: Pairwise Comparisons for estimated duration at CV2 across tonal contrasts (EH, H, M, L, and MR). The asterisk (*) shows the significant pairs.

To complement the statistical findings, Figure 4.14 presents a box plot illustrating the distribution of mean duration averaged across the five tonal categories at both the CV1 and CV2 positions. This visual summary reinforces the conclusion that CV2, responsible for lexical contrast, exhibits greater duration across all five tonal categories. In contrast, initial syllables (CV1) remain significantly shorter, with an average duration of approximately 100 ms across all tonal contrasts.

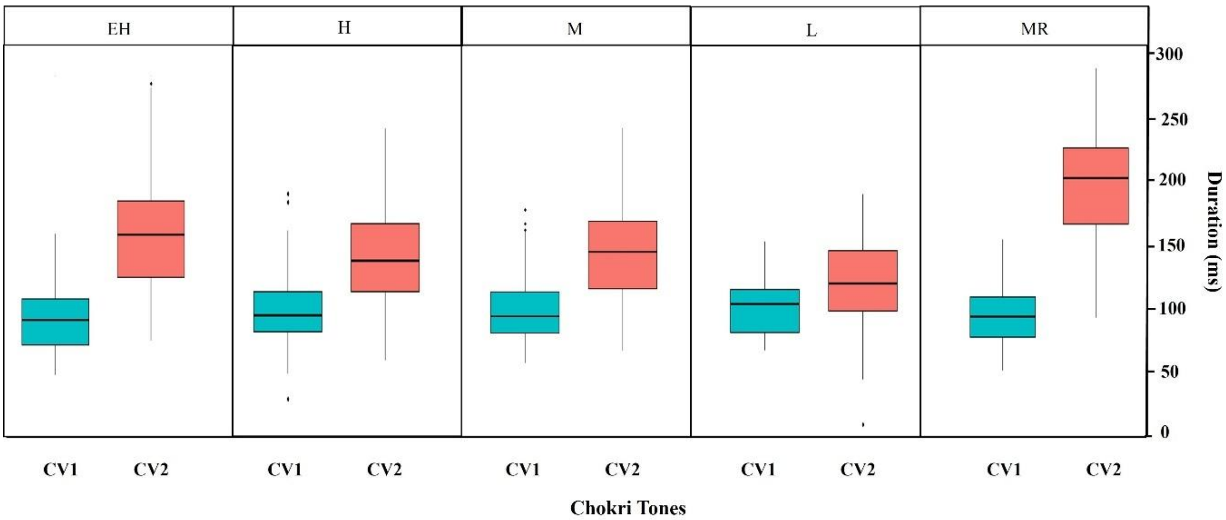


Figure 4.14: Mean duration (measured in ms), averaged across all the tokens produced by all the subjects for the five contrastive tones observed in bisyllabic words.

4.4.4.3. The Role of Intensity in Bisyllabic Words

Similar to f0 (Table 4.11) and duration (Table 4.13), the estimated mean intensity values at the CV1 position for the five tone exhibit relatively small differences, as shown in Table 4.15. All values fall within a narrow range of approximately 65.8 to 67.1 dB. The M tone has the highest estimated intensity at 67.1 dB, followed closely by the L tone at 67 dB. The EH and H tones are slightly lower, with means of 66.6 dB and 66.2 dB, respectively, while the MR tone has the lowest intensity at 65.8 dB. These numerical differences, at the 95% confidence intervals for all tones show considerable overlap, indicating that the observed variation is minimal and not statistically robust. The small standard errors further suggest that the intensity values are stable across tonal contrasts, but the overall range is too narrow to imply any meaningful acoustic differences at CV1. Given the lack of statistically significant differences in intensity at CV1 and the absence of meaningful tonal separation, intensity cannot be considered a reliable cue for distinguishing tonal contrasts at this syllable position. As such, pairwise comparisons were not conducted.

Tone	Estimated Intensity (dB)	SE	95% CI (Lower)	95% CI (Upper)
EH	66.6	0.311	66	67.2
H	66.2	0.302	65.6	66.8
M	67.1	0.31	66.5	67.7
L	67	0.303	66.4	67.6
MR	65.8	0.303	65.2	66.4

Table 4.15: Estimated mean values of Intensity (dB) at CV1for the five tones (EH, H, M, L, and MR), along with their standard errors (SE) and 95% confidence intervals (CI).

At the CV2 position, pairwise comparisons of estimated intensity values across the tonal contrasts, as shown in Table 4.16, reveal several large numerical differences, particularly involving the L tone. However, none of these differences reach statistical significance after applying correction for multiple comparisons. For example, the difference between the EH and L tones is 7.79 dB, and between H and L is 5.84 dB, both of which are substantial in magnitude, but neither is statistically significant. Similarly, the contrast between M and L shows a difference of 5.14 dB, and the difference between L and MR is -5.48 dB, but all of these comparisons fail to achieve statistical significance. While the L tone consistently shows lower intensity compared to the other tones, the overlapping confidence intervals and lack of statistically significant t-values suggest that these differences may reflect normal variation rather than systematic acoustic cues associated with tone. Furthermore, the contrasts among EH, H, M, and MR reveal relatively small differences in intensity, with no pair showing meaningful separation. In summary, although some tone pairs, especially those involving the L tone, exhibit notable numerical differences in intensity at CV2, the absence of statistically significant results suggests that intensity does not serve as a reliable acoustic marker for distinguishing tonal contrasts at CV2.

Contrast	Intensity (dB)	SE	t-value	p-value
EH-H	1.89	0.41	4.53	0.1
EH-M	2.68	0.42	6.33	0.1
EH-L	7.79	0.41	18.59	0.1
EH-MR	2.33	0.41	5.56	0.1
H-M	0.79	0.41	1.88	0.59
H-L	5.84	0.41	14.27	0.24
H-MR	0.43	0.41	1.05	1
M-L	5.14	0.41	12.19	0.65
M-MR	-0.35	0.41	-0.84	1
L-MR	-5.48	0.41	-13.20	0.1

Table 4.16: Pairwise Comparisons for estimated Intensity at CV2 across tonal contrasts (EH, H, M, L, and MR). The asterisk (*) shows the significant pairs.

To visually summarize these findings, Figure 4.15 presents a box plot of mean intensity values across the five contrastive tones at both the CV1 and CV2 positions. This visual representation reinforces the conclusion that intensity does not reliably differentiate tonal contrasts at either position. While there are some numerical differences, particularly involving the L tone at CV2, these differences do not achieve statistical significance.

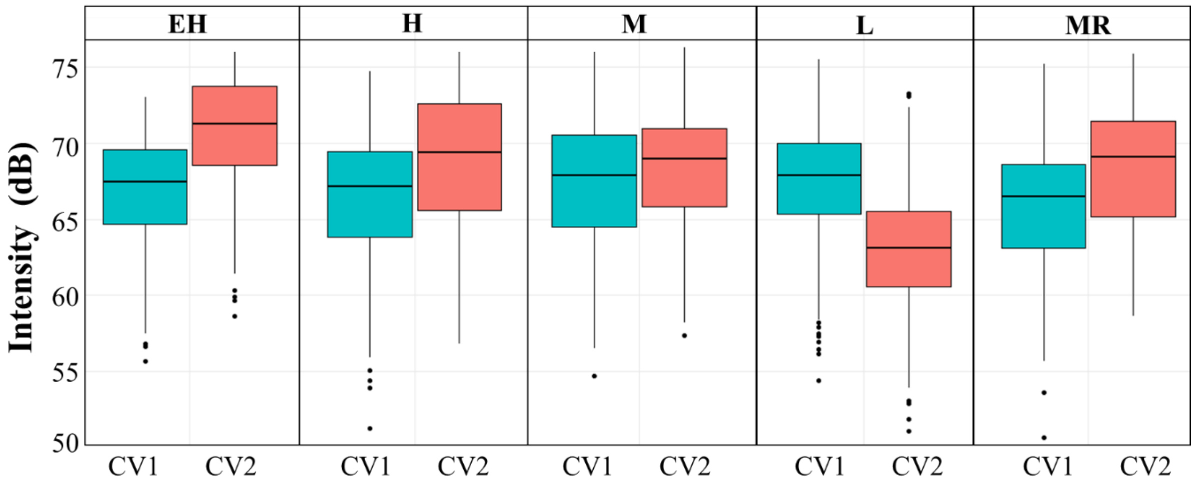


Figure 4.15: Mean intensity (dB), averaged across all the tokens produced by all the subjects for the five contrastive tones observed in bisyllabic words.

4.5. Perception of Lexical Tones

Tonal languages primarily rely on fundamental frequency differences to distinguish between contrastive tones. However, tone is not always solely dependent on f0 variations; other acoustic parameters, such as duration, intensity, and phonation cues (e.g., breathy, creaky, tense, and lax phonation), may also contribute to tonal contrasts in some languages (Yip, 2002). These additional cues can enhance tonal distinctiveness, particularly in cases where f0 differences are subtle especially in languages with large tonal inventory.

In certain tonal languages, phonation differences serve as primary or secondary cues for tone distinction. For example, languages such as Mandarin Chinese and Thai primarily rely on F0 contours, whereas languages like Burmese and Vietnamese incorporate phonation types (breathy vs. creaky) to reinforce tonal contrast (Brunelle, 2009). Languages can also use intensity and duration, particularly in distinguishing level and contour tones (Gandour, 1977).

Despite the potential role of multiple acoustic cues, preliminary observations from this study suggest that contrastive cues such as duration and intensity were not significantly utilized in Chokri. This suggests that f0 remains the primary distinguishing feature for tone identification in the target language. To explore this aspect, we designed a perception test to examine if the native speakers rely exclusively on f0 distinctions, i.e., whether or not the f0 fluctuations alone is enough to perceptually differentiate the contrastive tones by the native speaker.

4.5.1. Designing Perception Test

A perception test was designed to investigate if the native speakers rely solely on f0 distinctions when identifying contrastive tones. The test was based on a recorded mid-tone sound file, selected as the reference tone. We carefully selected four seperate mid tone words as our primary stimuli. The f0 of these reference words was synthesized using Praat to create systematic variations simulating different tonal categories, viz., EH, H, L and MR. The modifications were applied in two ways:

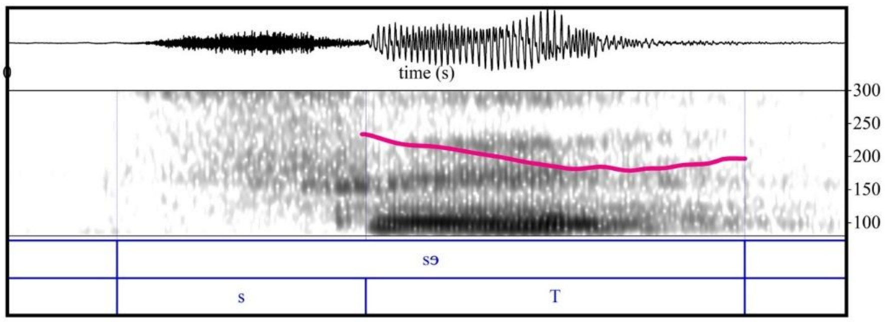
(i) Manipulation of Level Tones

The mid-tone served as the baseline reference. For higher tones, the f0 was increased by 20 Hz per step, and for lower tones, the f0 was decreased by 20 Hz per step. The following changes were applied:

- **+20 Hz** for the high tone (H)
- **+40 Hz** for the extra high tone (EH)
- **-20 Hz** for the low tone (L)

Figures 4.16 – 4.18 illustrate these modifications*:*

Original pitch track



Manipulated pitch track (+40 Hz)

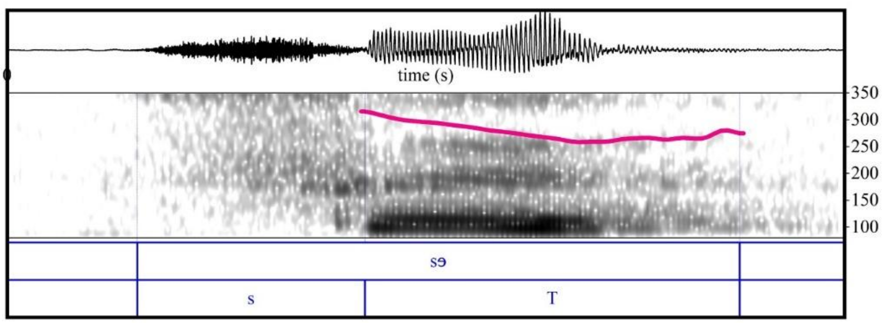


Figure 4.16: The original pitch track, with the mid-tone serving as the baseline reference (upper panel), and the modified pitch track increased by +40 Hz to create the extra high tone [EH] (lower panel).

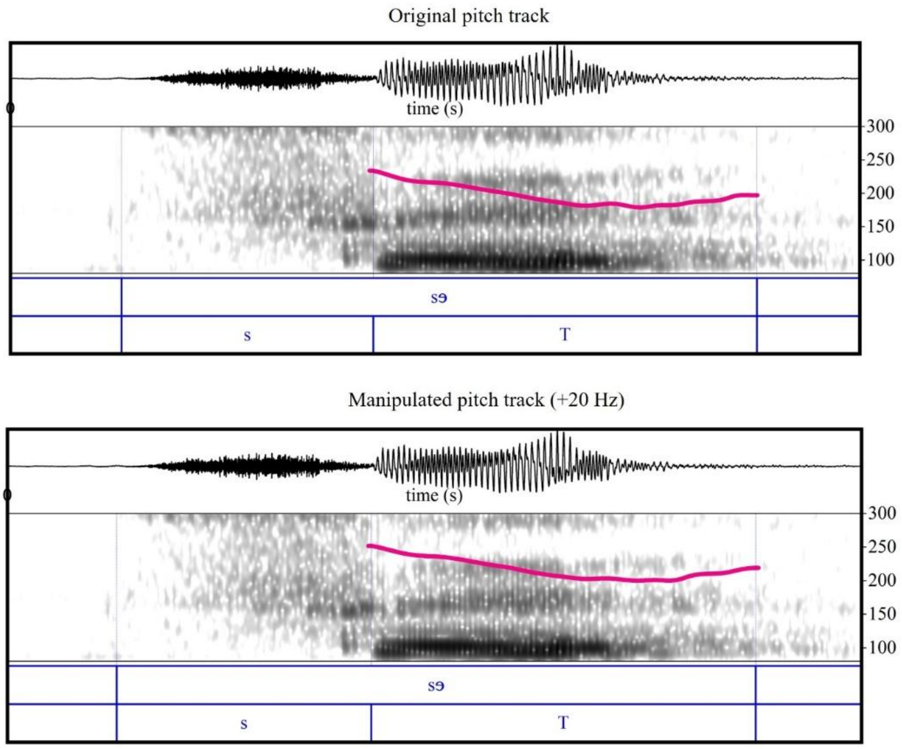


Figure 4.17 : The original pitch track, with the mid-tone serving as the baseline reference (upper panel), and the modified pitch track, increased by +20 Hz to create the high tone [H] (lower panel).

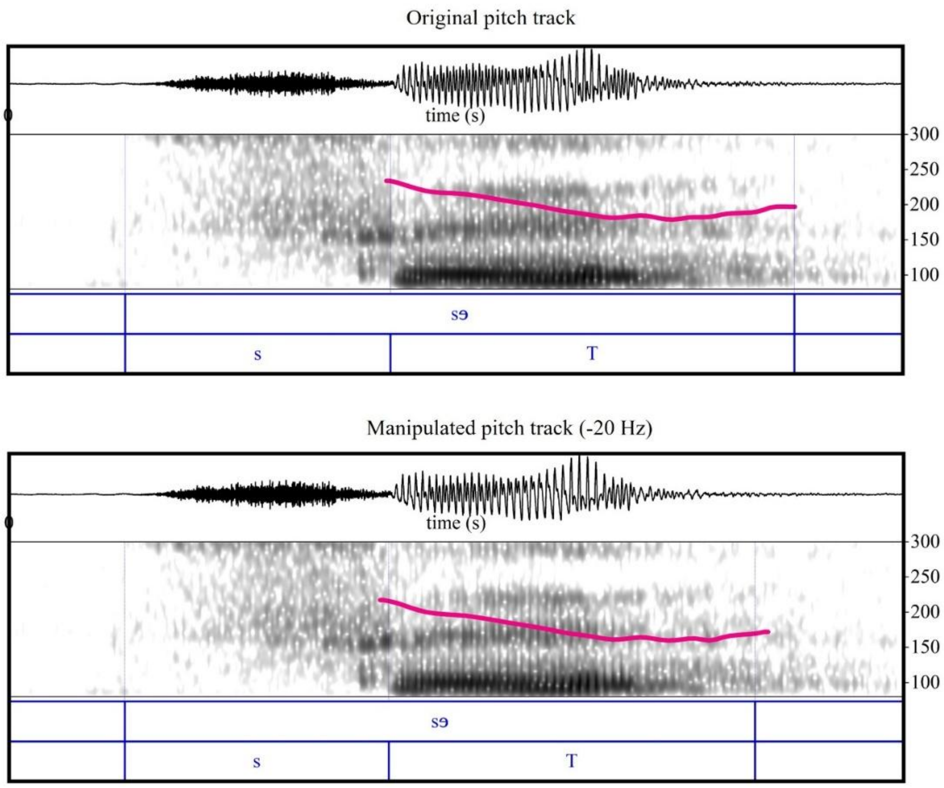


Figure 4.18: The original pitch track, with the mid-tone serving as the baseline reference (upper panel), and the modified pitch track, decreased by -20 Hz to create the low tone [L] (lower panel).

(ii) Manipulation of Contour Tones

For the creation of a mid-rising tone, a progressive pitch increase was applied. Specifically, the last 70% of the pitch track was raised by +25 Hz, generating a perceptible rising contour. Figure 4.19 illustrates this modification.

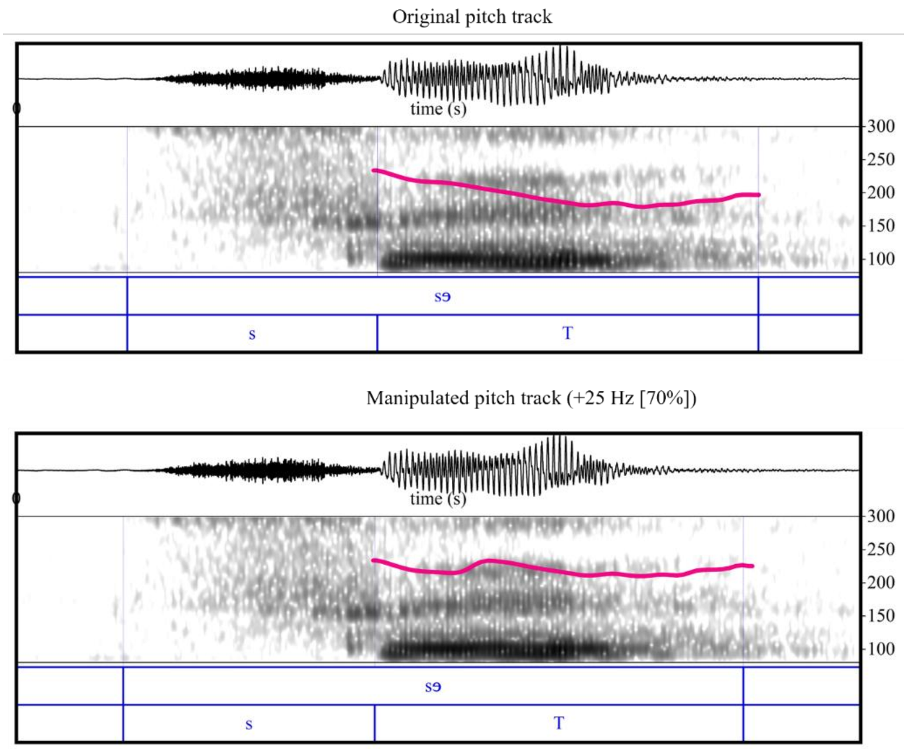


Figure 4.19: The original pitch track, with the mid-tone serving as the baseline reference (upper panel), and the modified pitch track, which has been increased by +25 Hz to 70% of the pitch contour to generate the mid-rising tone [MR] (lower panel).

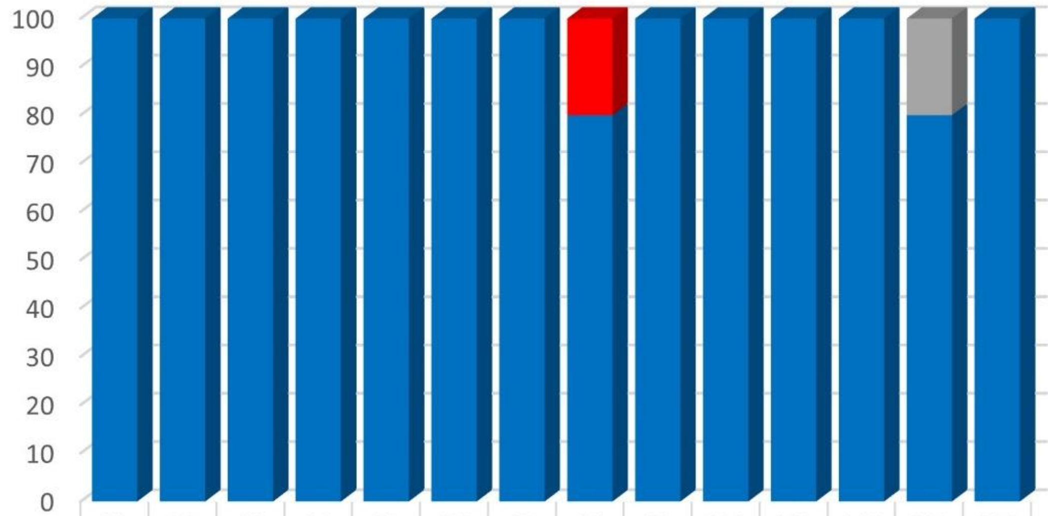
The test consisted of systematically manipulated stimuli where only f0 directionality was varied, while other acoustic properties (such as duration, and intensity) remained constant. Each stimulus was paired with six response options: the meaning of the target word, four contrastive meanings, and a "Not sure" option.

A total of 14 subjects (6 males and 8 females), aged between 20 and 60, participated in the experiment. The participants were instructed to listen to each manipulated audio clip up to a maximum of three times and select the option that best represented the meaning of the word they heard. Each stimulus was presented five times in a randomized order to minimize order effects. Each subject thus, listened to 25 tokens (5 words with 5 way tonal contrasts x five repetitions).

4.5.2. Results and Discussion

Stimulus 1:

Participants listened to the manipulated pitch stimuli presented in Figure 4.16, where the pitch track was adjusted to derive the extra high (EH) tone. The tone perception/identification results, shown in Figure 4.20, reveal that one participant misidentified tone 1 (EH tone) as tone 2 (H tone). Additionally, Speaker 13 marked one instance as “not sure.” However, all other participants correctly identified the tone in all repetitions. These results suggest that the EH tone is mostly perceptually distinct based on their f0 fluctuations alone.



	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
Not sure													20	
Tone 2								20						
Tone 1	100	100	100	100	100	100	100	80	100	100	100	100	80	100

Figure 4.20: Responses to stimulus 1 (EH tone), where mid-tone was systematically converted to EH. S= subject, Tone 1= (manipulated) EH tone (embedded as one of the six options), Tone 2= H tone (embedded as one of the six options). The Y-axis represents the response accuracy in percentage.

Stimulus 2:

For the high tone, participants listened to the manipulated stimulus presented in Figure 4.17, where the pitch track was adjusted accordingly. As reflected in Figure 4.21, all participants correctly identified tone 2 (high tone) in every instance, with no errors recorded. This suggests that tone 2 (H tone) is highly perceptually distinct and easily recognizable among native speakers.

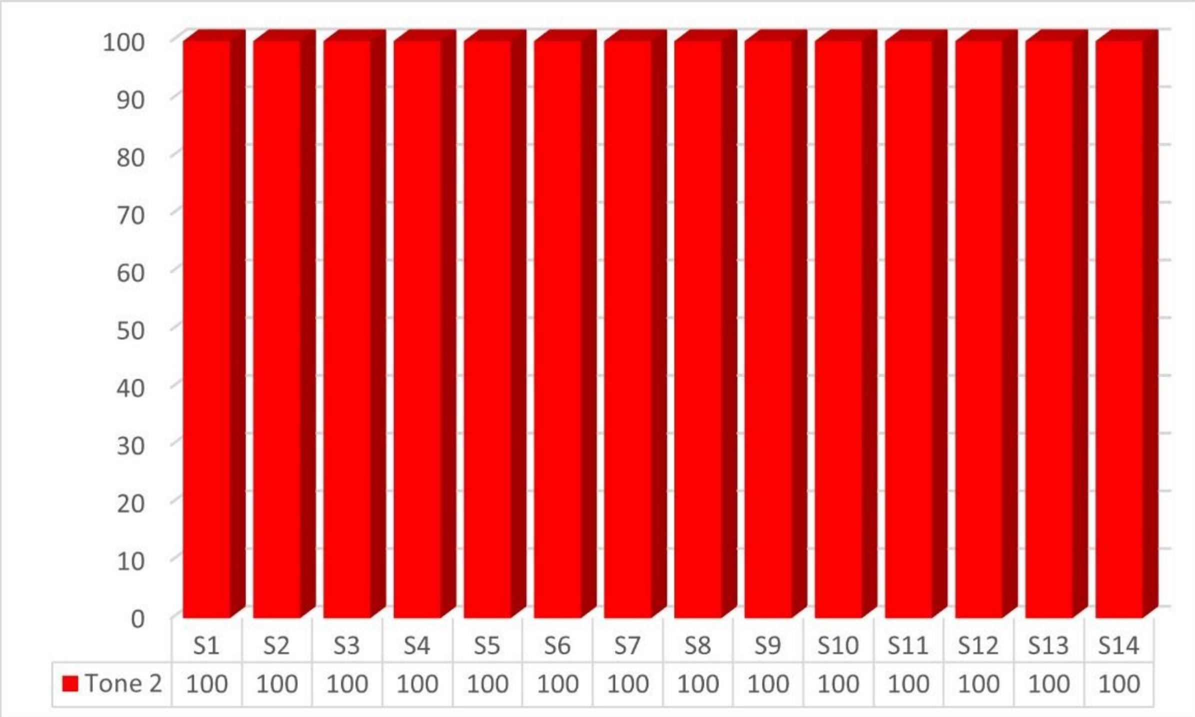
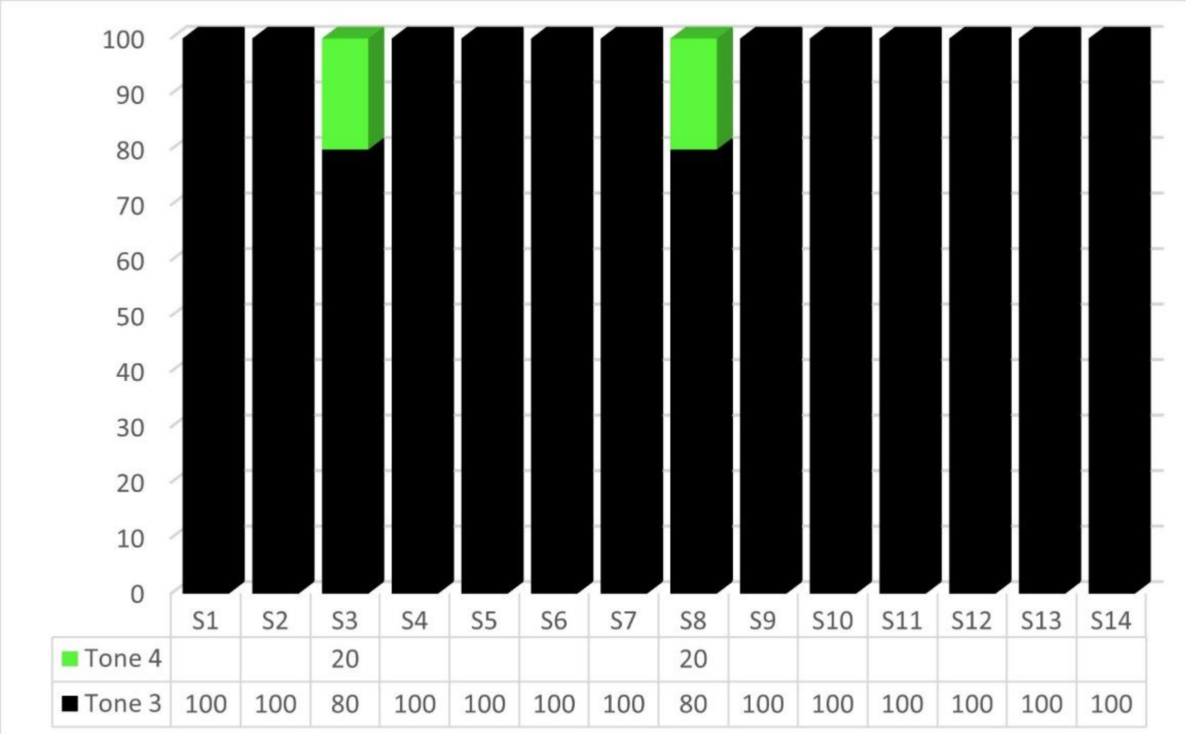


Figure 4.21: Responses to stimulus 2 (H tone), where mid-tone was systematically converted to H tone. S= subject. All the subjects predicted all the repetitions correctly for stimulus 2. S= subject. The Y-axis represents the response accuracy in percentage.

Stimulus 3:

For Tone 3, the original pitch track (the baseline reference) was used without any modification. This unaltered pitch served as the standard for comparison in the perception experiment. Responses to stimulus 3, shown in Figure 4.22, revealed that two participants misidentified Tone 3 (mid tone) as tone 2 (high tone) in one instance each. Despite these occasional misidentifications, the majority of participants accurately identified tone 3 (mid tone) across all trials, indicating that the baseline reference tone is mostly perceptually distinguishable.



	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
Tone 4			20					20						
Tone 3	100	100	80	100	100	100	100	80	100	100	100	100	100	100

Figure 4.22: Responses to stimulus 3 (mid tone), where the original sound found was used without any manipulation. S= subject, Tone 4= L tone (embedded as one of the six options), Tone 3= M tone (embedded as one of the six options). The Y-axis represents the response accuracy in percentage.

Stimulus 4:

Participants listened to the manipulated pitch stimuli presented in Figure 4.18, where the pitch track was adjusted to derive the low tone by lowering the f0 by 20 Hz. The results for stimulus 4, shown in Figure 4.23, were predominantly accurate, with 95% of total responses correctly identifying the tone. However, one participant (Subject 3) misidentified tone 4 (low tone) as tone 3 (mid tone) in a single instance. This indicates that the low tone is largely identifiable based on f0 distinction alone by the native speakers.

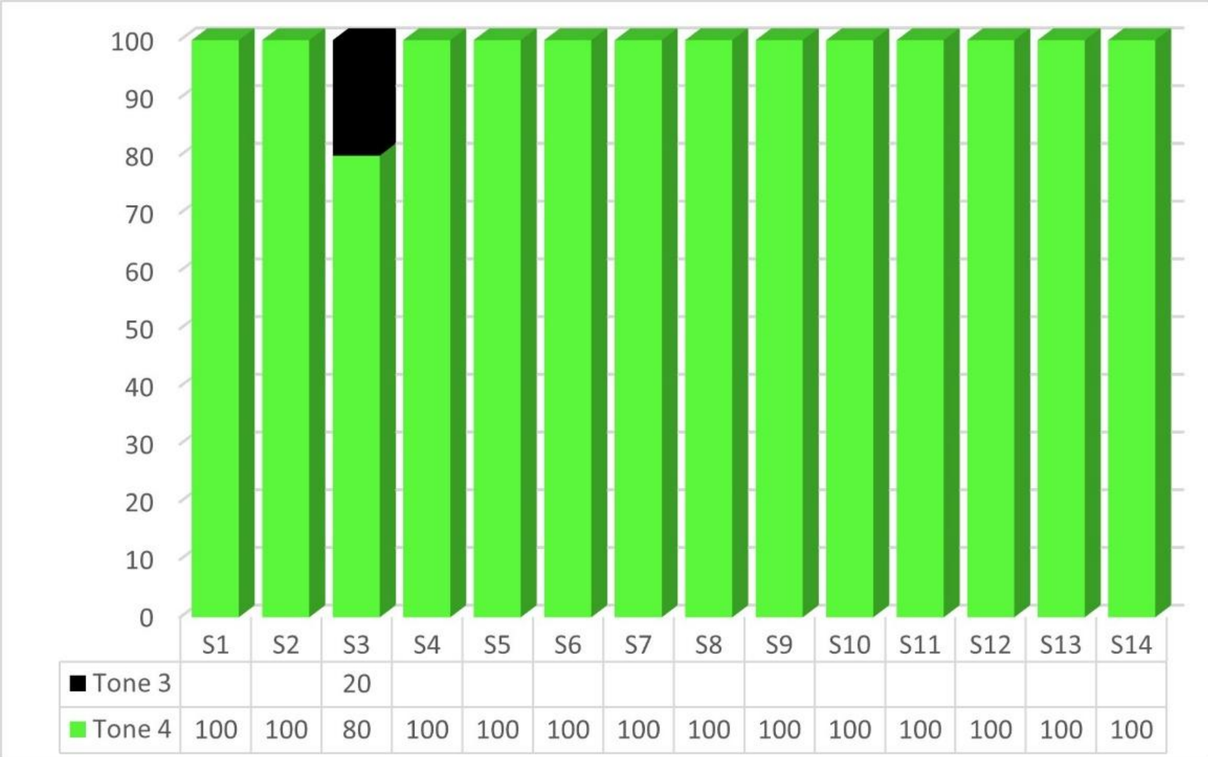


Figure 4.23*:* Responses to stimulus 4 (L tone), where mid-tone was systematically lowered to create the low tone stimulus. S= subject, Tone 3= mid tone (embedded as one of the six options), Tone 4= L tone (embedded as one of the six options). The Y-axis represents the response accuracy in percentage.

Stimulus 5:

For the mid-rising (MR) tone, participants listened to the manipulated pitch stimuli presented in Figure 4.19, where the pitch track was adjusted to create a rising contour. The responses to stimulus 5 (MR), as shown in Figure 4.24, indicate that all participants consistently identified the tone correctly across all instances. This suggests that the mid-rising tone is highly perceptually salient among native speakers, and its contour is easily discernible

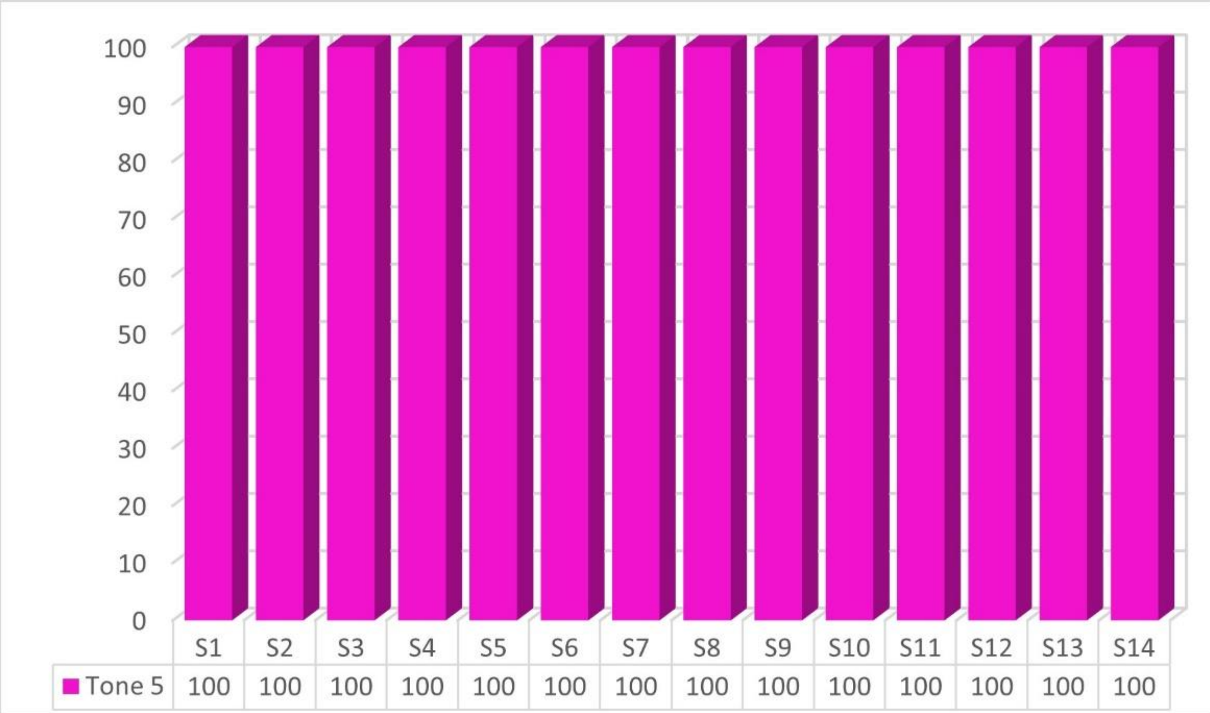


Figure 4.24*:* Responses to stimulus 5 (MR tone), where mid-tone was systematically adjusted to create the MR stimulus. S= subject, All the subjects predicted all the tokens correctly for stimulus 5. The Y-axis represents the response accuracy in percentage.

Figure 4.25 presents a summary of all the responses across all five tonal categories. The data reveal that tones 2 (H) and 5 (MR) were identified with complete accuracy, while minor misidentifications occurred with tones 1 (EH), 3 (M), and 4 (L). Confusion was observed between tones 3 (M) and 2 (H), as well as between tones 4 (L) and 3 (M) in one occurrence each. Despite these occasional errors, the overall results indicate a high level of accuracy in tone identification based solely on f0, reinforcing its role as the primary perceptual cue in the language. Notably, it must be considered that, the mid-rising tone, by default carry slightly higher durational values compared to the level tones as observed earlier. In the perception study, while the mid-tone was used to create the MR stimulus, we did not extend the duration that would have otherwise accompanied with a contour MR tone. However, as we have observed, the synthesized MR stimulus was correctly predicted across all the repetitions by all the speakers. This indicate that the significant durational differences observed between the contour MR and the four level tones (viz., EH, H, M, and L) could be a phonetic bi-product and may not be required to perceptually differentiations of contrastive tones by the native speakers.

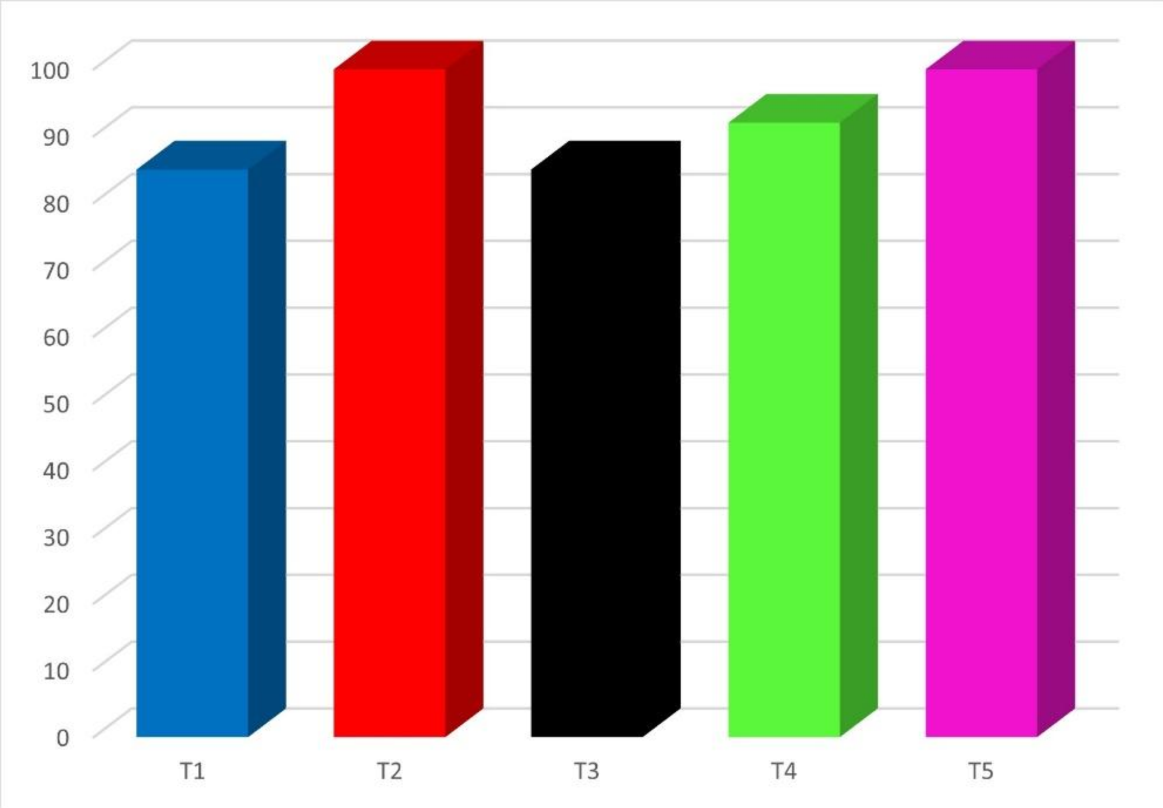


Figure 4.25*:* Overall response to all the stimuli by all the Subjects. T1=EH, T2=H, T3=M, T4=L, and T5=MR. The Y-axis represents the response accuracy in percentage.

4.6. Tone Bearing Unit in Chokri

Gussenhoven (2005) defines Tone Bearing unit (TBU) as "the element in the segmental structure to which tone associates." This concept is integral to understanding the relationship between tone and segmental elements in a language. According to Yip (2002), tone exists independently of segmental and prosodic elements, similar to a musical melody. It is realized at the surface level only when it is linked to a specific segment or prosodic unit, such as a syllable or mora, on which it is pronounced. This linking process determines the TBU for a language.

The TBU can vary depending on the language's phonological structure. In languages with monomoraic, open CV syllables, where each syllable carries exactly one tone, the TBU may be the vowel, mora, or syllable. Yip (2002) further explains that in cases where a language includes both light and heavy syllables, differing in the number of tones they can carry, the TBU is typically the mora. However, in languages where both light and heavy syllables can bear the same number of tones, the TBU is considered to be the syllable

As discussed in Chapter 2 (Section 2.2.3), Chokri does not distinguish between long and short vowels. Moreover, as noted in Chapter 3 (Section 3.2), Chokri adheres to a strict open syllabic structure, restricting the occurrence of codas. Therefore, every syllable in Chokri presents a light monomoraic CV or CCV structure. In this structure, each syllable is typically one mora in length and carries exactly one tone.

However, during instances of speech simplification, segments may be deleted while preserving the tone of the deleted segment. In such cases, the tone of the deleted segment attaches to a host syllable. This process often results in the vowel of the hosting syllable undergoing lengthening to accommodate the tone, which is necessary for maintaining tonal integrity. This lengthening transforms the light monomoraic syllable (bearing a single mora) into a heavy bimoraic syllable (bearing two moras). In this transformation, each mora becomes a carrier of tone, allowing the tonal pattern to be fully realized (more on this is discussed in Chapter 5).

This suggests that in Chokri, the TBU is the mora, rather than the vowel or the syllable. This is further illustrated in Figure 4.26, which represents how the tone is linked to the mora in Chokri, particularly when vowel lengthening occurs to accommodate the tone. The figure demonstrates how the syllable structure and tonal pattern interact, reinforcing the idea that in Chokri, the mora functions as the TBU.

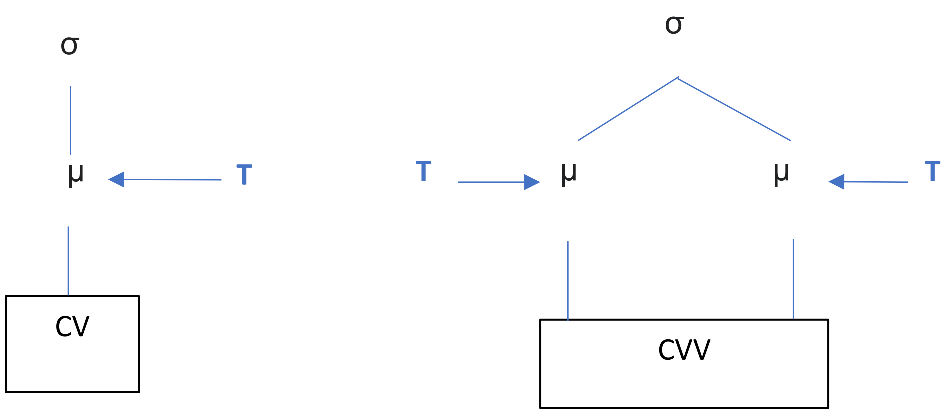


Figure 4.26: Representation of TBU in Chokri

This figure visually demonstrates how the vowel lengthening process in response to tonal demands results in a bimoraic structure. It also highlights how the mora acts as the TBU in Chokri, making the tonal pattern fully realized across syllables that undergo this transformation.

4.7. Conclusion

Studies indicates that tone production is influenced by articulatory constraints (Xu, 1994, 2004), which inherently limit the number of contrastive level tones that can exist within a language. The posited maximum number of level tone is five, and to date, no language has been documented with more than five contrastive level tones. Maddieson (1978) observed that approximately half of the tonal languages sampled exhibit two-way tonal contrasts, while languages with four or five contrastive level tones are relatively rare. Attributing to the articulatory limitations of pitch production and perception, languages with huge inventory of contrastive level tones are disfavoured as maintaining large inventory of contrastive level tones in speech communication is challenging (Kuang, 2013).

Chokri is one of the rare languages hosting four contrastive level tones within the favourable pitch range of 100 Hz between the EH and L tones by both the genders (see Figures 4.3 and 4.6). The f0 differences observed between the level tones in Chokri surpass the *Just-Noticeable Difference* (JND) threshold of around 9 Hz reported by Silverman (2003), thus ensuring that the tones remain distinct for native speakers. Kuang (2013) further suggested that pitch difference of at least 20 Hz is necessary for maintaining phonological contrast, even in languages with large tonal inventories. In case of Chokri level tones, we observed that this threshold too is well maintained in this language.

Interestingly, Chokri's tonal system deviates from prior literature on high-tone duration, as observed by Faytak and Yu (2011), who suggested that high tones typically exhibit shorter durations. In Chokri, however, the high tone (H) has a longer duration than the low tone (L), though this difference is not statistically significant. Our study also confirms that Chokri's four level tones remain static in terms of f0 fluctuations (see Figures 4.5 and 4.6), with their duration also consistent across the tonal range (see Figure 4.7). This finding supports the view that Chokri's level tones are genuinely level rather than being phonetically dynamic.

The acoustic characteristics of the mid-rising (MR) contour tone in Chokri reveal a pitch movement from mid-level to high-level, creating what can be described as a 'glide' or 'cluster tone' (Pike, 1948). This contour tone is significantly longer in duration compared to the level tones, adding a temporal distinction in addition to its f0 variation. This further highlights the unique nature of the tonal system in Chokri.

The results from the perception test reinforce the conclusions drawn from the production experiment. Native speakers were able to identify the tones accurately based on (synthesized) f0 fluctuations alone, with minimal errors. Notably, there was no significant gender-based variation in tone perception, which emphasizes the robustness of f0 as the primary cue for tonal identification in Chokri.

Chokri's tonal system is a rare and complex structure, featuring multiple contrastive level tones and a unique contour tone. The findings from both the production experiment and perceptual analysis suggest that Chokri speakers rely predominantly on f0 cues for tonal distinction. The distribution of tonal properties observed in the bisyllabic words further highlights the importance of the final syllable in carrying the tonal contrast. In the forthcoming chapter, we will discuss in detail the grammatical functions of tone observed in this language.

References

Alexander, J. A. (2010). *The Theory of Adaptive Dispersion and Acoustic-Phonetic Properties of Cross-Language Lexical-tone Systems* (Doctoral dissertation), Northwestern University, Evanston).

Baken, R. J., and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*. Singular Publishing Group.

Bao, Z. (1999). *The Structure of Tone*. Oxford University Press.

Bauer, R. S., and Benedict, P. K. (1997). *Modern Cantonese Phonology*. Mouton de Gruyter.

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

Brunelle, M. (2009). Tone Perception in Northern and Southern Vietnamese. *Journal of Phonetics,* 37(1), 79-96.

Chao, Y. (1930). *A System of Tone Letters*. Le Maître Phonétique, 45, 24-27.

Clements, G. N., and Ford, K. C. (1979). Kikuyu Tone Shift and Its Synchronic Consequences. *Linguistic Inquiry, 10*(2), 179-210.

Connell, B. (2000). The Perception of Lexical Tone in Mambila. *Language and Speech*, 43(2), 163-182.

Duanmu, S. (2007). *The Phonology of Standard Chinese* (2nd ed.). Oxford University Press.

Faytak, M., and Yu, A. C. L. (2011). Tonal Dynamics and Duration in Level-Tone Languages. *Proceedings of the 17th International Congress of Phonetic Sciences*, 667–670.

Frazier, M. 2009. Tonal Dialects and Consonant-Pitch Interaction in Yucatec Maya. *New Perspectives in Mayan Linguistics,* WPLMIT 59, 59-82.

Gandour, J. (1977). On the Interaction Between Tone and Vowel Length: Evidence from Thai Dialects. *Phonetica,* 34(1), 54-65.

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

Gope, A., Pal, A., Tetseo, S., Gogoi, T., Joanna, J., & Borah, D. (2024). Multi-class Identification of Tonal Contrasts in Chokri Using Supervised Machine Learning Algorithms. *Humanities and Social Sciences Communications,* https://www.nature.com/articles/s41599-024-03113-2

Hyman, L. (2007). Kuki-Thaadow: An African Tone System in Southeast Asia. UC Berkeley *Phonology Lab Annual Reports*. 3. 10.5070/P77G41N67R.

Hyman, L.M. (2011). Tone, is it different?. *The handbook of Phonological Theory.* Blackwell Publishing.

Hyman, L. M. (2007). Universals of Tone Rules: 30 Years Later. In T. Riad and C. Gussenhoven (Eds.), *Tones and tunes: Studies in phonetics and phonology* (Vol. 1, pp. 1–34). Mouton de Gruyter.

Kuang, J. (2013). The Tonal Space of Contrastive Five Level Tones. *Phonetica*. 70(1-2):1-23. doi: 10.1159/000353853. Epub 2013 Oct 23. PMID: 24157433.

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

Liljencrants, J., and Lindblom, B. (1972). Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language*, 48(4), 839–862.

Maddieson, I. (1978). Universals of Tone. In J. H. Greenberg, C. Ferguson, and E. A. Moravcsik (Eds.), *Universals of human language* (pp. 335–365). Stanford University Press.

Maddieson, I. (n.d.). Tone Systems of the World's Languages [Abstract]. Retrieved from chrome-Extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.eva.mpg.de/lingua/conference/2012_Tone/pdf/abtsract_Maddieson.pdf

Merrifield, W., Edmondson, J. 1999. Palantla Chinantec: Phonetic Experiments on Nasalization, Stress, and Tone. *International Journal of American Linguistics*, *65*(3). http://www.jstor.org/stable/1265786

Pike, Kenneth L. (1948). *Tone languages: A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion.* Ann Arbor: University of Michigan Press.

Pulleyblank, D. (1986). *Tone in Lexical Phonology*. https://dspace.mit.edu/handle/1721.1/15508

Shue, Y.-L., Keating, P., Vicenik, C., and Yu, K. (2011). VoiceSauce: A Program for Voice Analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)* (pp. 1846–1849).

Silverman, K. (2003). Pitch Discrimination During Speech Perception. *In Proceedings of the 15th International Congress of Phonetic Sciences (*ICPhS), 655–658.

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

Sonderegger, M. (2023). *Regression Modeling for Linguistic Data*. United Kingdom: MIT Press.

Tetseo, S., and Gope, A. (2021). Tonal Contrast and Grammatical Tone in Chokri. [Paper presentation]. 43rd International Conference of Linguistic Society of India, CIIL-Mysore.

Xu, Y. (1994). Production and Perception of Coarticulated Tones. *Journal of the Acoustical Society of America* 95: 2240-2253

Xu, Y. (1997). Contextual Tonal Variations in Mandarin. *Journal of Phonetics* 25: 61-83

Xu, Y. (2004). Understanding Tone from the Perspective of Production and Perception. *Language and Linguistics* 5: 757-797. Author version

Yip, M. (2002). *Tone.* Cambridge Press

Zhang, J. (2002). *The Effects of Duration and Sonority on Countour Tone Distribution: A Typological Survey and Formal Analysis* (1st ed.). Routledge. https://doi.org/10.4324/9781315024134