Chapter 3: Downtrends

3.1. General Overview of Downtrend phenomena

One of the major areas of research in works on intonation in tonal and non-tonal languages is downtrend patterns, i.e., the gradual lowering of f0 from the starting towards the end of utterances. The presence of a downtrend in a language implies that the pitch of a tone occurring later in the utterance will be lower than that of the same tone occurring at the beginning. Although debatable, such progressive pitch descent is often attributed to phonetic effects (Connell, 2001). Downtrends include (i) declination- the time-dependent lowering of f0; (ii) downstep- the lowering of f0 of H tones induced by the intervention of an L tone; and (iii) final lowering (Gussenhoven, 2004.) In intonation-only languages, earlier studies on downtrends were focused chiefly on declination (Maeda, 1976; Pierrehumbert, 1979; Cooper and Sorensen, 1981). However, downtrend analysis in many African tone languages reports the presence of downstep, a process involving H and L tones occurring alternatively, wherein the later H in an HLH sequence has a lower f0 due to the influence of the intervening L tone (Connell, 2001). Downtrend analysis sheds light on the phonetics and phonology of pitch realization in a given language. Apart from providing insights into the general intonational pitch patterns at utterance levels, downtrend analysis is also relevant due to its interaction with other intonational properties like focus and phrase boundary. In Mandarin, for instance, the rate of declination is higher after narrow focus elements than that of plain utterances (Shih, 1998). Downstep, on the other hand, is weakened by on-focus f0 raising and post-focus-compression (Xu, 1999; Wang et al., 2022). In Tswana, downstep is blocked by phonological phrase boundaries (Zerbian and Kügler, 2015, 2021). Ishihara (2007) has shown that the downstep is weakened by a strong phrase boundary and a focused H tone after the L tone. In many languages, pitch reset at new phrase boundaries that disrupt the continuing downward slope of f0 is a common boundary marking device. This phenomenon has been reported in languages like Bemba (Kula and Hamann, 2017), Japanese (Kawahara and Shinya, 2008), Ékegusií (Hieber, 2016), and Chichewa (Downing, 2017), where pitch reset marks domain edges. Suspension of downstep is often employed as a phonological means of marking question sentences in many languages, including Danish (Gussenhoven, 2004) and Twi (Hyman, 2001).

Declination refers to the global downward trend of f0 contour stretched over phrases or utterances (Stewart, 1983; Ladd, 1984; Liberman and Pierrehumbert, 1984; Pierrehumbert and Beckman, 1988). It is the "gradual modification (throughout a phrase or utterance) of the phonetic backdrop against which the phonologically specified f0 targets are scaled" (Connell and Ladd, 1990). Generally considered a phonetic effect, declination is considered universal in declarative utterances across languages. It is the most striking ‘global attribute’ in utterances in both English and Dutch (t’ Hart et al., 1990). Declination affects IP level f0 in Japanese (Beckman and Pierrehumbert, 1986; Deguchi and Kitagawa, 2002; Pierrehumbert and Beckman, 1988). In Cantonese, both utterance declination and phrase declination simultaneously affect sentence level f0 realization in mixed tone and identical tone sequences (Flynn, 2003; Fox et al., 2008).

There is a rich body of work that explores the physiological accounts of this downtrend pattern, which attributes its occurrence to falling sub-glottal pressure and laryngeal configurations (Liberman, 1967; Titze, 1989; Strik and Boves, 1995). Maeda's (1976) '*Tracheal Pull Hypothesis*' holds that the sternum is lowered when the lung volume decreases. The larynx is physically connected to the sternum. It gradually goes down during the production of an utterance, resulting in a declining f0. Ohala and Ewan (1973) state that the laryngeal combinations that are required to produce a higher f0 are more complex than those needed to produce a fall. This means that speakers ought to favour making smaller rises instead of declines. Additionally, listeners would anticipate a f0 drop in utterances because of the f0 downward trending tendency. It would also become "a variable that speakers use with communicative intent" due to their usage of it as a perceptual cue for domain borders (Ohala, 1981). According to Xu (1997), rather than being a basic principle of intonation, declination is the sum total of various linguistic factors and local physiological constraints.

Although the earlier works on declination were conducted in non-tonal languages (Liberman and Pierrehumbert, 1984, for English; Appels, 1985, for Dutch), there have been challenges in accurately identifying the true characteristics of declination. The declining pitch slope in these languages has to be estimated from sparsely situated located f0 peaks or valleys. Moreover, as the syllables are not specified for tone, an observable pitch contour is often the outcome of many factors (Shih, 1998.) The accent-to-accent decay observed in non-tonal languages is deemed to be a phonological downstep, termed catathesis, instead of time-dependent lowering (Pierrehumbert and Beckman, 1988). In this context, tone languages offer scope for a more accurate understanding of declination. While the declination effect can be seen for all tones, in tone languages, they are clearer through analysis of sentences containing words with the same lexical tones, like all H or all L sequences. Declination is primarily present in declarative utterances. However, many researchers observed the absence of declination in some tonal languages, which has been one of the exciting findings from the works on intonational downtrends. For example, declination and final lowering are not found in Yoruba for sentences with all-H and all-M tones (Connell and Ladd, 1990). In Mambila, Tone 1, the highest tone, shows little tendency for downtrend (Connell, 2001).

Another type of downtrend frequently reported in African tone languages concerns alternate H and L tone sequences. H tone following an L tone is realized with a lowered f0 compared to the H tone preceding an L tone. This process is called automatic downstep or downdrift (Stewart, 1983). Non-automatic downstep, on the other hand, is the lowering of an H tone without a conditioning L tone being physically present. A floating L tone triggers such a downtrend. One of the characterizing features of a downtrend is that a downstepped H tone resets the pitch ceiling for the H tones that will subsequently occur. Moreover, downstep is cumulative as a successive series of downstep results in a progressively lowering pitch track (Connel, 2001). According to Rialland (1997), downstep has three distinct properties: i) downstep does not affect lexical tonal specifications of tones; i.e., a high tone never changes to lower tones due to downstep b. The effect of downstep extends to the entire tonal sequence in its domain. And, c. The phonetic realization of downstep varies from language to language, and it is defined differently by different authors. Languages like Igbo (Laniran, 1992) and Bimoba (Snider, 1998) show the presence of both downdrift and non-automatic downstep. However, tone languages may avoid the downstep effects of either type in order to preserve the lexical tone specifications of the syllables. Such a language is Mambila, which shows no instance of downstep in sequences of its four tones (Connel, 1999).

Downstep is a property of intonational pitch accents, too. The phonological nature of such downstep has been proven in several languages, including American English (Liberman and Pierrehumbert,1984), Japanese (Beckman and Pierrehumbert, 1986), and

Catalan (Estebas-Vilaplana, 2000). Liberman and Pierrehumbert (1984) showed that the downstep of H tones in the HLH sequence in American English can be characterized by a constant rate of change from one peak to another regardless of utterance length and number of intervening syllables.

Final lowering is the more abrupt and excessive lowering of f0 at the end of speech domains like phrases or utterances (Gussenhoven, 2004; Liberman and Pierrehumbert, 1984; Truckenbrodt, 2004). Bolinger (1978) and Liberman (1982) considered final lowering to be a property of most of the languages. Over the years, the presence of final lowering has been reported in English (Liberman and Pierrehumbert, 1984), Japanese (Pierrehumbert and Beckman, 1988), Dutch (Gussenhoven and Rietveld, 1988), Kipare (Herman, 1996), and Yoruba (Connell and Ladd, 1990; Laniran, 1990). Welmers (1973) reports the presence of final lowering in several discrete-level tone languages of Africa. Gussenhoven (2004) opines that final lowering is grammaticalized in many languages where it encodes a post-lexical meaning, i.e., the finality of the utterance.

As indicated in Chapter 2, pitch accents in Sylheti and lexical tones in Chokri both exhibit downward f0 movement within the IPs. In this chapter, we explored the types of downtrends present in the languages and their precise nature through statistical and mathematical analysis. Through this study, we seek to answer the following research questions:

i) Given the presence of phrasal tones constituting APs in Sylheti, how is the surface realization of these tones influenced by the f0 downtrend that spans over the IPs? Do lexical tones impact the downstepping of f0 peaks of phrasal tones?

ii) Given Chokri's intricate nature of lexical tones (five-way tonal contrasts), and somewhat simple two-way tonal contrasts in Sylheti, how much are these tones affected by post-lexical downtrends of f0?

iii) What types of downtrends manifest in Sylheti and Chokri? How do we quantitatively evaluate the presence of the downtrend effects in the languages?

iv) What is the precise nature of downward f0 slopes resulting from downtrend effects, and how can these be modeled effectively?

3.2. Experimental Design and Procedure

3.2.1. Participants and Dataset

Five native Sylheti (three male, two female) speakers and five native Chokri (three male, two female) participated in the recording of speech data for downtrend analysis. All the Sylheti speakers are from the Dharamnagar region of Tripura and are aged between 18 to 33. The Chokri participants, aged between 18 and 55, were residents of Thipuzu village in the Phek district of Nagaland. The data recording task in both the languages involved scripted sentences on a computer screen, with participants instructed to produce them naturally. The entire datasets were repeated five times to ensure variability, incorporating significant intervals between repetitions. The datasets for this experiment were designed with the assistance of a native speaker and cross-verified with two other native informants to ensure the authenticity of the language variety under consideration. The details of the dataset can be found in the Appendix section.

The dataset used for Sylheti consists of 11 scripted neutral declarative sentences. These sentences vary in length (8 syllables, 9 syllables, 10 syllables, and 11 syllables) and in the number of f0 peaks. Various sequences of lexical tonal contrast were included in the study to account for the possible impact of lexical tones in the downtrend of surface f0 contours.

The scripted sentences utilized in the study of the Chokri downtrend comprised four sets, each containing four distinct types of sentences. The dataset's structure employed for the experiment is outlined as follows:

1a. All high-tone (H) sequences;

1b. All low-tone (L) sequences;

1c. All mid-tone (M) sequences; and

2. H tone sequences with an intervening L: HLHH, HHLH, and HHLHH

It must be noted that not all the tone sequences (or intervening tonal sequences) form (semantically) meaningful sentence constructions in Chokri. Therefore, this study concentrates only on the tonal sequences that form meaningful and natural sentences in

Chokri while preparing the dataset (The complete datasets are provided in Appendix I and II).

3.2.2 Data Recording and Annotation

Recordings were digitized at a sampling frequency of 44.1 kHz and 32-bit resolution. For detailed analysis, individual sentence files were manually annotated and labeled at the syllable level using *Praat* (Boersma and Weenink, 2012). Syllable-wise mean f0 and time-normalized f0 values were extracted using *ProsodyPro* (Xu, 2013). The resulting pitch contours are graphically represented as line charts for visual inspection. This systematic grouping allows for a clear and structured analysis of the f0 patterns in the sentences.

3.2.3 Statistical and Mathematical Modeling

A one-way repeated measures analysis of variance (RM ANOVA) was performed to assess the statistical significance of observed differences in fundamental frequency (f0) using R (version 4.2.3) (R Core Team et al., 2013). The analysis utilized the function *aov(response(f0) ~ factor(time) + Error(factor(syllable)), data = df*) (Sonderegger, 2023) in R. Both mathematical and statistical modeling are incorporated to investigate the phenomena of declination and downstep. Fitted lines (Linear fits for Sylheti final lowering and Chokri sentences with all H and L tone sentences) and exponential decay fits (for Chokri All M tone sequences) were utilized for modelling downward slopes of f0. The mathematical equations were fitted using *Origin* software (version 8.1) and in *Python* using the *linregress* function from the *scipy.stats* package. To assess the accuracy of the prediction of downstep models, the *r2_score* function from *sklearn.metrics* in Python is used to compute the R-squared value.

It was noted that the range of f0 varied between male and female speakers, primarily due to differences in vocal cord size across genders. To account for potential intra-speaker and inter-speaker anomalies or hidden covariates across all tokens produced by different subjects, a linear mixed-effects regression was conducted in R using the *lme4* package (Bates et al., 2009). The *lm(y ~ x)* function, representing a basic linear model, was employed, where "y" represents the dependent variable (f0) and "x" signifies the independent variable (time points). Considering the potential influence of syllable count, the model was further adjusted as *lm(f0 ~ time + syllable + gender)*. The *lm()* model was preferred over a mixed-effects model *[lmer()]* since the primary focus of this study was to examine the overall effects of time, syllable count, and gender on f0 across all speakers. While mixed-effects models are valuable for accounting for random effects due to speaker-specific variability, this study specifically aimed to investigate the fixed effects of the independent variables, viz., time and syllable count, while considering gender as a fixed effect on f0. Additionally, our dataset did not demonstrate significant variability attributable to individual speaker differences that would require the inclusion of random effects.

3.3. Results and Discussions

3.3.1. Downtrend in Sylheti

The presence of a downstep in a language implies that consecutive f0 peaks in utterances will be scaled with gradually lowering f0 values. In Sylheti, neutral declaratives are divided into APs marked by an L* pitch accent and Ha boundary tones. To analyze the nature of downstep in the language, the peaks created by Ha tones were investigated. While producing the data used for this study, the speakers tended to divide IPs into two or three APs. This less crowding of AP peaks is the result of recursive APs that enclose lexical elements along with their surrounding functional constituents (see Chapter 2, section 2.3.3).

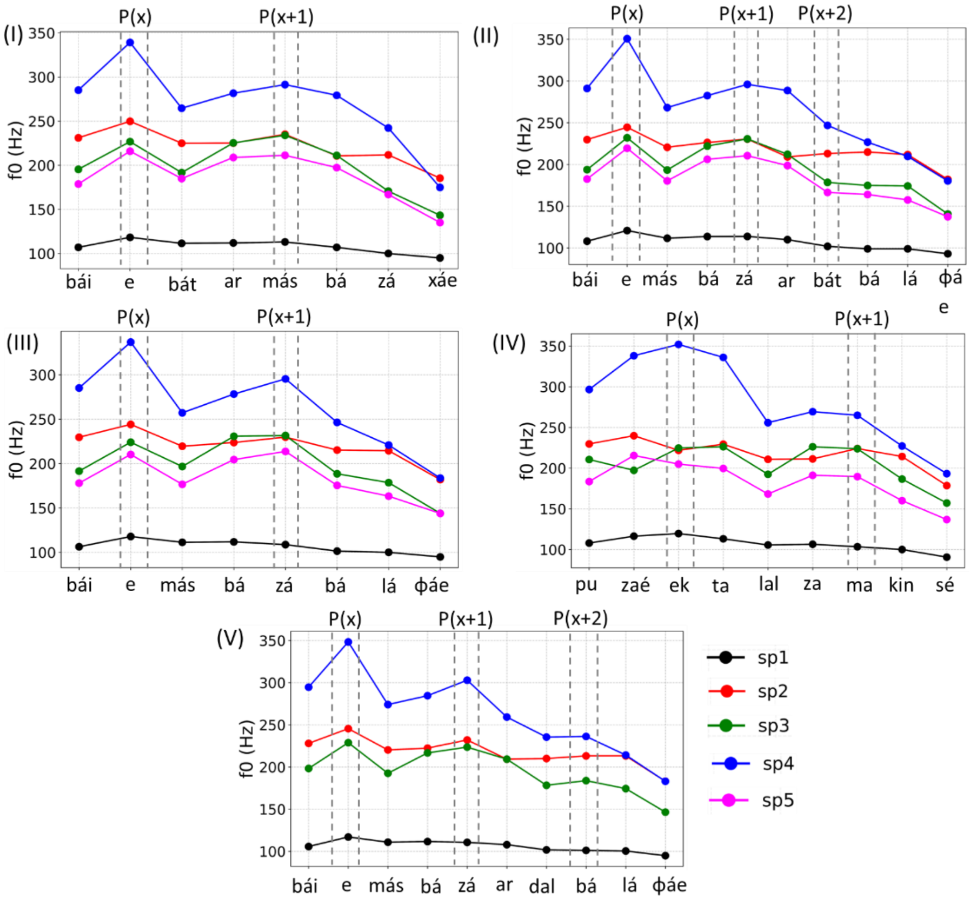

Figure 3.1: The average f0 contours of Sylheti for five speakers, represented by different colors: black for sp1, red for sp2, green for sp3, blue for sp4, and magenta for sp5 across different sentences: (I) [bái-e bát ar más bázá xáe] '(my) brother eats rice and fried fish,' (II) [bái-e más bázá ar bát xáe], 'my) brother eats fish fry and rice' (III) [báie más bázá bálá ɸáe] '(my) brother likes fried fish '(IV) [puzaé ekʈa lal zama kinsé], 'Puja bought a red dress,',' (V) [bái-e más bázá ar dail bálá ɸái] '(my) brother likes fried fish and lentil'. The dashed vertical lines indicate the positions of the peaks, denoted as P (x), which represents the peak height at position x within each sentence.

From the visual inspection of the f0 contours in **Figure 3.1 (I-V),** it is evident that the female speakers show higher peak values across all sentences and an overall higher pitch range when compared to the male speakers. Despite the variations in pitch ranges among the speakers, there is a noticeable consistency in the contour shapes within individual speakers across different sentences. This consistency suggests that each speaker maintains the characteristic pitch pattern of repeated APs regardless of the sentence length. There are two or three peaks for the sentences, with P(x) indicating the height of the peak at position x in each sentence. The visual inspection of the peaks overall suggests that the first peak (P(x)) is often the highest across all speakers, followed by subsequent peaks (P(x+1) and P(x+2)) that tend to decrease in height. It is to be noted that the underlying lexical tones do not impact the f0 peaks, as they change the scaling of the L* pitch accents, influencing the interpolation lines. Moreover, scaling changes induced by IP medial H tone roots do not alter the specification of the phrasal tones; they change the height of the pitch line interpolating L and Hs.[1]

3.3.2. Sylheti Final lowering:

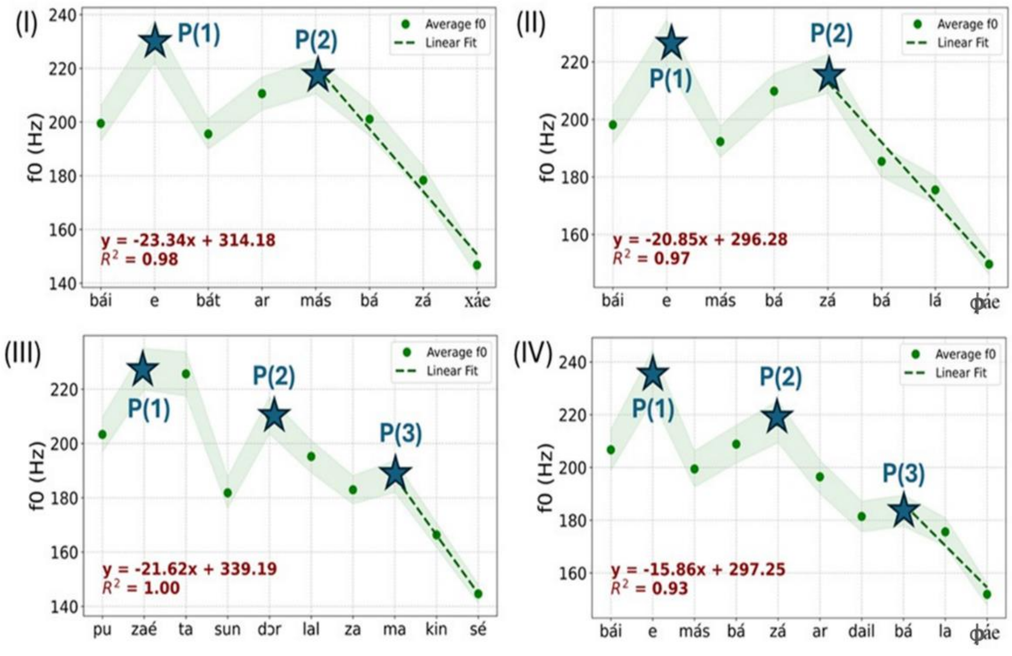

Figure 3.2: The average f0 contours for different sentences, displayed in panels (I-IV). The dashed lines represent the linear fits of the f0 contours, displaying the final lowering. Peaks, denoted as P(1), P(2), and P(3), are highlighted using stars.

Figure 3.2 (I-IV) presents a detailed view of the average f0 contours for different sentences, highlighting the presence of peaks. The prominent peaks, denoted as P(1), P(2), and P(3), are marked with stars, representing the highest tonal points in the pitch contour of each sentence. To investigate the phenomenon of final lowering in the Sylheti language, we focused on the final f0 points and applied linear fits to these data. The linear equation used is y =m * x+c, where m is the slope, indicating the rate of change of f0 with respect to the syllable, and c is the intercept, representing the offset value. These linear fits were implemented in *Python* using the *linregress* function from the *scipy.stats* package. The linear fits offer a simplified model of the f0 trajectory, which helps to illustrate the overall trends in pitch movement throughout the sentence. A consistent observation across all sentences is that the slope (m) is negative, indicating that the f0 decreases as it moves towards the end of the sentence. This negative slope provides

[1] The pitch tracks and spectrograms are provided in Appendix III

sufficient evidence of the downtrend phenomenon in Sylheti, where the pitch naturally declines as it progresses towards the end of sentences.

In our study, the *r2_score* function from *sklearn.metrics* in Python is used to compute the R-squared value, which is a statistical measure of how well the regression predictions approximate the real (actual observed) data points. It ranges from 0 to 1, with 1 indicating that the model explains all the variability in the response data around its mean. The high R^2 values, all above 0.9, further confirm that the linear model accounts for a substantial portion of the variability in the data. This suggests that, for these sentences, pitch movement generally follows a predictable pattern of final lowering. The slope values range between 15-23 Hz per syllable, demonstrating that this pattern of lowering is consistent regardless of whether the sentences contain two or three peaks.

3.3.3 Phonological nature of Downstep of f0 peaks in Sylheti:

The downward f0 movement, as seen in peak-by-peak descent, is considered to be phonological instead of a phonetic byproduct of the articulatory process. The constant and predictable rate of change in f0 serves as motivation for establishing downstep as phonological events by Liberman and Pierrehumbert (1984) This entails that the sentence length or number of syllables between peaks does not affect the f0 lowering rate. As seen in **Figures 3.1** and **3.2**, the initial f0 and final f0 of utterances with different lengths are almost the same. Similarly, the pitch excursions from the beginning of the utterances to the first peak are also in the same range. This hints at the likelihood of a downtrend of peak height to be a phonological property instead of being a phonetic declination.

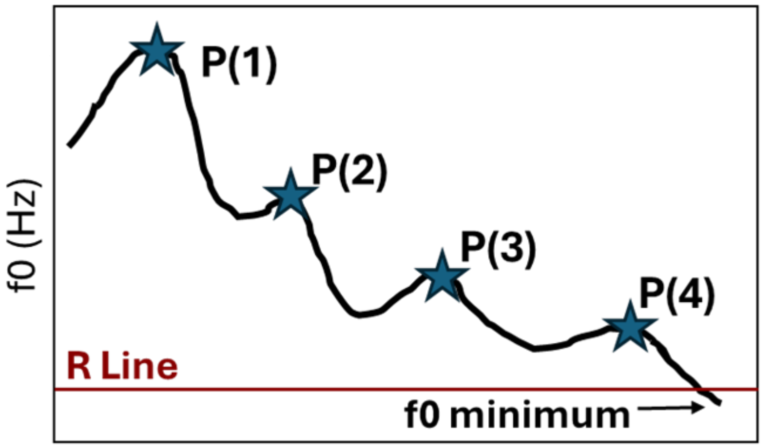

Figure 3.3: Adaptation of Liberman and Pierrehumbert's (1984) fit as an exponential fit to a constant nonzero asymptote, where peaks are highlighted using stars, and the R line stands for the reference line.

Liberman and Pierrehumbert (1984) suggested that in English downstepping contours, the height of each peak can be predicted by applying a consistent reduction to the previous peak's f0 value. The decay between peaks is calculated as a proportion of the second peak relative to the first, adjusted above the speaker's baseline level.

Figure 3.3 shows the adaptation of Liberman and Pierrehumbert's (1984) fit as an exponential fit, where peaks are highlighted using stars, and the R line stands for the reference line (R). Liberman and Pierrehumbert introduced the concept of the R in the mathematical model fit. The decay ratio between consecutive peaks, known as the downstep constant, is calculated using the following equation:

Downstep ratio (r) = (P(x+1) - R) / (P(x) - R)

where P(x) represents the peak height at position x, and R is the reference line value. For each speaker, the reference line R is calculated as:

R = (Mean f0_last peak + Mean f0_minimum) / 2

where Mean f0_last peak is the average f0 value of the last peak across all sentences, and Mean f0_minimum is the average minimum f0 value across all sentences. This equation calculates R as the midpoint between the mean f0 of the last peak and the mean minimum f0 across all sentences.

Once these values are obtained, the f0 height of any given peak can be calculated or predicted as a constant fraction of the previous peak's height, scaled relative to the reference line of each speaker. The following equation expresses this relationship:

P(x+1)_predicted = R + r * (P(x) - R)

where P(x) represents the peak height at position x, r is the downstep ratio, and R is the reference line.

This equation shows that each peak height can be predicted by applying a scaling factor (the downstep ratio) to the difference between the previous peak height and the reference line, then adding the reference line value back to the scaled result.

Application of this predictive calculation to our data confirms that when the f0 contour includes only two peaks, the downstep ratio effectively describes the relationship between these peaks. The downstep ratio (r) [the second peak [P(2) to the first peak [P(1)] is found to be 0.69±0.09, represented by the average value ± standard deviation. This ratio remains consistent, indicating a stable relationship between consecutive peaks, regardless of the temporal location of the second peak. For instance, even if the number of intervening syllables varies, as shown in **Figure 3.1(I-V),** the downstep ratio remains relatively unaffected. This stability suggests that the downstep ratio is a reliable metric for capturing the pattern of pitch reduction between peaks. The consistency of r across different contexts implies that the relationship it describes is a fundamental characteristic of the speech pattern rather than being influenced by the precise timing or location of the peaks.

However, in utterances with more than 2 peaks, the downstep ratio (r) [the third peak [P(3) to the second peak [P(2)] is found to be 0.27 ±0.17, which is also the last peak of the sentence. The fact that the standard deviation is almost as large as the average value indicates significant variability, suggesting that this downstep ratio is not stable and is subject to considerable fluctuations. Such variability implies that the transition from the second to the third peak is less predictable and may be influenced by other factors, the final lowering in particular. It is widely observed across many languages that the final peak in an f0 contour exhibits a more pronounced f0 descent. This phenomenon has been documented in languages such as American English (Liberman and Pierrehumbert, 1984) and Mexican Spanish (Prieto et al., 1996). Along with a movement towards a f0 minimum, the final lowering can also be characterized by a sharper drop in pitch at the last peak of an utterance, which falls below the levels predicted by the standard downstep rule. This indicates that the f0 value of the last peak cannot be modeled using the same downstep ratio applied to non-final peaks. To account for this, Liberman and Pierrehumbert (1984) suggested modeling the final peaks of an utterance using a lowering constant (l). This constant represents a fraction of the peak value predicted by the downstep rule positioned above the reference line. The f0 value of the last peak is calculated using the following formula:

P = R + l * (P(down) - R)

where P is the height of the last peak, P(down) is the peak height predicted by the downstep rule, R is the reference line, and l is the final lowering constant. The ratio determines the value of l:

l = (P(obs) - R) / (P(down) - R)

where P(obs) is the observed peak height. Once the l is computed, we can use it to predict the next peak's height using the formula:

P(x+2)_predicted = R + l* (P(x+1) - R)

where the P(x+2)_predicted is the predicted f0 contour.

The observed (P(x+2)_observed) and predicted values (P(x+2)_predicted) in Sylheti are prepared as numpy arrays in Python. Scatter plots were created to visualize the relationship between observed and predicted values. The R-squared scores were calculated to quantify how well the predicted values match the observed values. This methodology allows visual and quantitative evaluation of the performance of this predictive model in Sylheti as it combines statistical analysis with visual representation to provide a comprehensive understanding of the model's behavior.

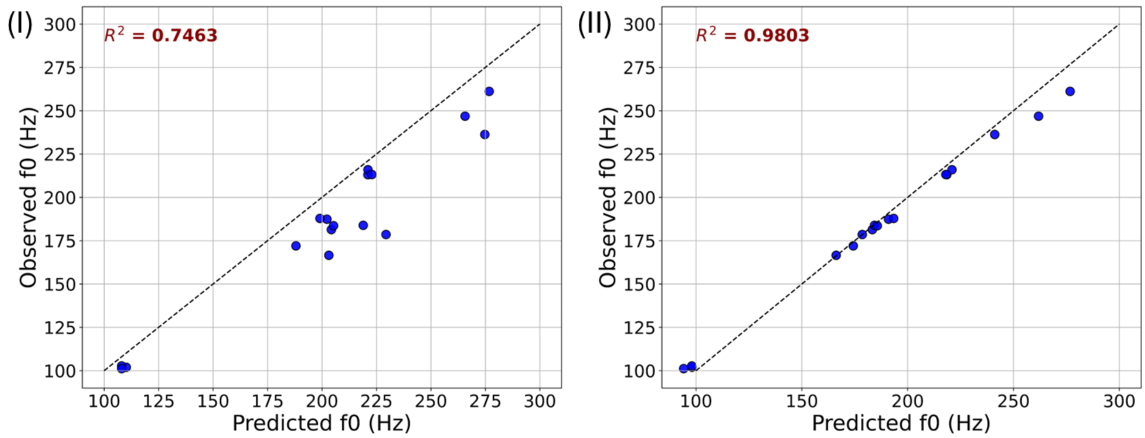

Figure 3.4: Comparison of predicted and observed f0 values for a given peak (P3). In [I], the predicted f0 values are calculated using the downstep ratio (r), while in [II], they are based on the final lowering constant (l). The dotted lines represent the regression lines, and the R^2 values are displayed to indicate the fit quality.

Figure 3.4(I-II) provides a comparison between the predicted and observed f0 values for a given peak (P3). In **Figure 4(I),** the predicted f0 values are calculated using the downstep ratio (r), and in **Figure 4(II)**, they are based on the final lowering constant (l). It demonstrates that the model is quite successful in predicting the f0 height of a given peak. This is evident from the fact that most of the observed f0 values are closely aligned with the predicted values, clustering around the x = y line. The high R^2 value of 0.98, obtained using the final lowering constant, demonstrates a very strong correlation between the predicted and observed f0 values. This suggests that using the final lowering constant provides a more accurate prediction of f0 values than the downstep ratio method, which achieved a lower R^2 value of 0.75 [see **Figure 3.4(I-II**)]. Our results thus show that downstepped f0 peaks in Sylheti can be mathematically modeled, affirming their phonological nature. Moreover, it highlights that in the presence of more than two peaks in Sylheti neutral declaratives, the final peak is influenced by the final lowering of f0 and can be successfully modeled using a lowering constant.

3.3.4. Declination in Chokri: Visual interpretation and statistical analysis

The results of the production experiment were designed to explore the declination trends in Chokri across all H, L, and M tone sequences, which indicate the presence of a phonetic declination process in the language for all three tone sequences. However, the rate of decline varies across the tone sequences. Time-normalized average f0 values, calculated for each male and female speaker, revealed that declination is most prominent in the initial syllable of the utterance. The first pitch drop occurs during the production of the first syllable, with the declination rate gradually decreasing as the utterance progresses. Despite the identical phonological specification of all syllables (high), the phonetic f0 value of late-occurring syllables is lower compared to those at the beginning.[2]

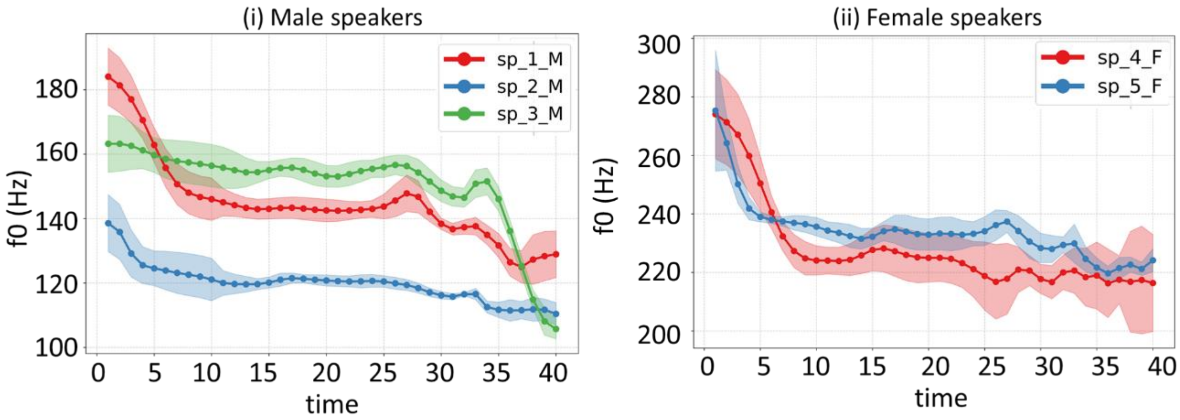

Figure 3.5: Time-normalized raw f0 contours of all H tone sequences, averaged across all the repetitions produced by each speaker in the utterance [í lɛ́ ʧɔ́ vá] "I am cooking"- (i) shows data from 3 male speakers and; (ii) shows data from 2 female speakers. The legends, [sp] stands for a subject, followed by the number of the subject and gender information, where M stands for male speaker, and F indicates a female speaker. The shaded error bars indicate the standard error of the aggregated data.

[2] The pitch tracks and spectrograms are provided in Appendix III

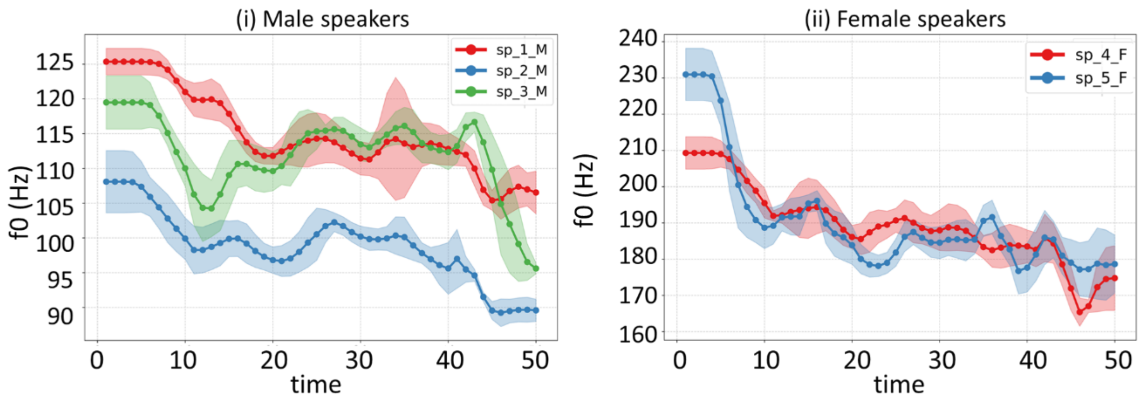

Figure 3.6: Time-normalized raw f0 contours of all L tone sequences, averaged across all the repetitions produced by each speaker in the utterance, [tʰì hù bə̀ tì tɔ̀] "(I) will boil some meat and eat"- (i) shows data from 3 male speakers and; (ii) shows data from 2 female speakers. The legends, [sp] stands for a subject, followed by the number of the subject and gender information, where M stands for male speaker, and F indicates a female speaker. The shaded error bars indicate the standard error of the aggregated data.

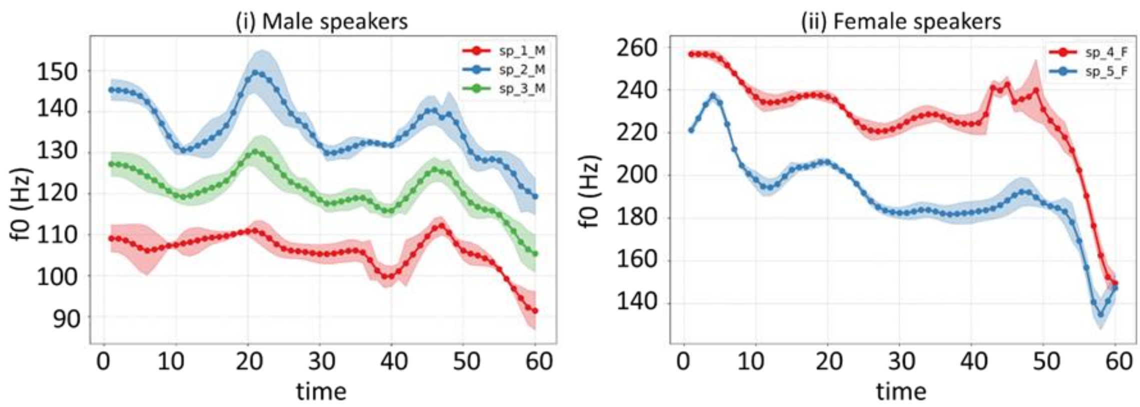

Figure 3.7: Time-normalized raw f0 contours of all M tone sequences, averaged across all the repetitions produced by each speaker in the utterance, [hīhī ā-lɛ̄ kʰɔ̄ lē] "this is my bag""- (i) shows data from 3 male speakers and; (ii) shows data from 2 female speakers. The legends, [sp] stands for speaker, followed by the number of subjects and gender information, where M stands for male speaker, and F indicates a female speaker. The shaded error bars indicate the standard error of the aggregated data.

Figures 3.5-3.7(i-ii) illustrate the time-normalized average f0 trends for male (i) and female (ii) speakers across all H, L, and M tone sequences. The standard error of the aggregated data is indicated with the shaded bars. A visual examination of the pitch tracks across the time-normalized f0 contours for different speakers and tonal sequences (H, M, and L) reveals that the f0 range for females is higher than that of males. The response of f0 as a function of time demonstrates a predominantly linear pattern for H and L tone sequences [see **Figures 3.5-3.6 (i-ii)**]. On the contrary, a curvature is observed for the M tone sequence [see **Figure 3.7 (i-ii)**].

The final lowering in the sentences with tonal sequences of all H and L tone sequences is not observed across the speakers' repetitions considered in this study, as depicted in **Figures 3.5-3.6(i-ii)**. Notably, these exhibit a distinctive decreasing pattern in the f0 range towards the end of the normalized time for one male speaker [sp_3_M (in green color)] across both H and L tone sequences. However, the qualitative analysis failed to classify this trend as a definite "final lowering accurately." It is essential to emphasize that this specific trend observed in a single male speaker's data is not entirely evident in the data of the female speakers or the two other male speakers. In contrast, the f0 trend observed in sentences with the M tone sequences displays a distinct decrease in f0 values towards the end of the normalized time, irrespective of gender. This observation hints at a final lowering trend in Chokri, as illustrated in **Figure 3.7(i-ii)**.

To rigorously assess the significance of the declination patterns illustrated in **Figures 3.5-3.7(i-ii)**, encompassing all H, L, and M tone sequences, a one-way repeated measures analysis of variance (RM ANOVA) was conducted. The results confirm that the f0 differences observed in each syllable are indeed statistically significant for the sequences of all H ($F[3, 12] = 66.7$, $p < 0.0001$*), all L ($F[3, 12] = 69.52$, $p < 0.0001$*), and all M ($F[5, 20] = 30.83$, $p < 0.0001$*) tones, respectively. Subsequent post hoc Bonferroni tests further affirm that the sequences of all H, M, and L tones are significantly different for each syllable. The f0 range of all H tones is higher than M and L. Similarly, the f0 range of all M tones is higher than L and lower than H tones. While this statistical analysis offers insights into group differences, it falls short in providing information on the intrinsic patterns or the nature of these tone sequences.

Statistical Modeling

The primary objective of our statistical modeling is to confirm our visual assessments and gain a deeper understanding of the inherent patterns and characteristics in the sequences of all H, M, and L tones. From a statistical modeling perspective, the choice between additive and interactive models emerges when examining these tone sequences, as elucidated in (Scott et. al., 1999) to examine these tone sequences. However, this study opts for an additive modeling approach to understand the impact of each predictor (gender, syllable, and time) on the response variable (f0 for various tone sequences of all H, M, and L tones). These predictors, being independent, facilitate deriving distinct estimations for the effect of each predictor. Given that the f0 range for female speakers is higher than their male counterparts, the primary focus lies in examining individual effects rather than interactive effects. Consequently, the predictor "gender" is intentionally excluded as an interaction term in the modeling adopted in this study.

Table 3.1: Statistical model output of f0 (Hz) for different tonal sequences - Response ∼ time + syllable (S_k) + gender.

Tone	Effects	Estimate	Std. Error	t value	Pr(>|t|)
H	(Intercept)	240.4563	1.4465	166.233	<0.0001*
	time	-1.1408	0.1663	-6.858	<0.0001*
	syllable (S_2)	4.7152	2.1431	2.200	0.02*
	syllable (S_3)	15.3410	3.5907	4.272	<0.0001*
	syllable (S_4)	15.3114	5.1698	2.962	0.003*
	gender M	-89.7632	0.9752	-92.043	<0.0001*
L	(Intercept)	204.53248	1.14108	179.245	<0.0001*
	time	-1.17328	0.13495	-8.694	<0.0001*
	syllable (S_2)	-0.08319	1.69491	-0.049	0.96
	syllable (S_3)	9.60382	2.91739	3.292	0.0015*
	syllable (S_4)	20.15289	4.23875	4.754	<0.0001*
	syllable (S_5)	29.40886	5.63614	5.218	<0.0001*
	gender M	-80.13828	0.79119	-101.288	<0.0001*
M	(Intercept)	236.3181	3.7665	62.742	<0.0001*
	time	-1.2861	0.3924	-3.277	<0.0001*
	syllable (S_2)	-0.4197	5.5358	-0.076	0.93
	syllable (S_3)	10.4171	8.7660	1.188	0.23
	syllable (S_4)	17.0594	12.4032	1.375	0.17
	syllable (S_5)	35.3654	16.1751	2.186	0.02*
	syllable (S_6)	27.8734	20.0057	1.393	0.16
	gender M	-89.6498	2.3008	-38.965	<0.0001*

The results of the generalized statistical modeling are presented in Table 1, with the response variable (f0 in Hz) based on various predictor variables (time, syllable, and gender) for different tone sequences (H, L, and M). A smaller p-value (< 0.05) indicates statistical significance, and the coefficient estimates (β) provide insights into the strength and direction of the relationships. The intercept, although lacking meaningful interpretation in this context, represents the estimated value of the response variable (f0) when all other predictor variables (time, syllable (S_k), and gender) are set to zero. However, setting these to zero is not meaningful in this scenario as the f0 depends on time and gender. Also, the syllable (S_k) is assigned by grouping the normalized time so that $S_k \in [1 + 10(k - 1), 10k]$, for k = 1, 2, 3, 4, 5, and 6. Therefore, $S_1 = 1$ when the normalized time is from 1 to 10, $\boldsymbol{S_2}$ = 2 when the normalized time is from 11 to 20, and so on. In contrast, the coefficient (β) for time represents the change in the response variable (f0) for a one-unit increase in the time variable while holding all other predictors constant. For H, L, and M tone sequences, the coefficients are -1.1408 Hz, -1.17328 Hz, and -1.4334 Hz, indicating a decrease per unit increase in time, respectively. The "gender" variable (Male or Female) demonstrates coefficients representing the difference in f0 between the two gender groups, with gender M as the reference group.

The coefficients for $\boldsymbol{S_k}$ with k ∈ [2, 6] denote the change in f0 for the corresponding syllables compared to the reference syllable (S1). For example, in the H tone sequences, $\boldsymbol{S_k}$ with k ∈[2, 4] have a significant positive effect of 4.7152 Hz, 15.34 Hz, and 15.31 Hz compared to the reference, respectively, with $p < 0.05$. A similar pattern is observed for the L tone sequences for $\boldsymbol{S_k}$, with k [2, 5]. Nevertheless, significant syllable predictors are notably lacking in the M tone sequences, as evidenced by p-values exceeding the statistical significance threshold ($p > 0.05$) for most syllables in our analysis. Furthermore, for the H and L tone sequences ($F[5, 794] = 1744$, and $F[6, 693] = 1767$, with $p < 0.0001$), the adjusted R squared values are 0.92 and 0.93, respectively.

In contrast, for the M tone sequence ($F[7, 292] = 231.6$, with $p < 0.0001$), the adjusted R-squared value is 0.84. The adjusted R-squared value indicates how well the linear regression model fits the data, reflecting the proportion of variance in f0 explained by the predictor variables (time, syllable, and gender). For the H and L tone sequences, the statistical model accounts for 93% of the variance in f0. This suggests that the predictor variables (time, syllable, and gender) collectively exhibit a robust linear explanatory effect on f0. In contrast, the adjusted R-squared value for the M tone sequence is 0.84, signifying that the statistical model explains only 84% of the variance in f0. This is notably less than the explanatory power observed for the H and L tone sequences, indicating a relatively weaker linear association with f0 in the M tone sequences, with less than a 10% reduction in variance compared to H and L tone sequences. This indicates that the linear trends of the H and L tonal sequences are better than the M tonal sequence, suggesting a potential differentiating pattern in the latter.

3.3.4.1. Mathematical Modeling of Chokri Declination

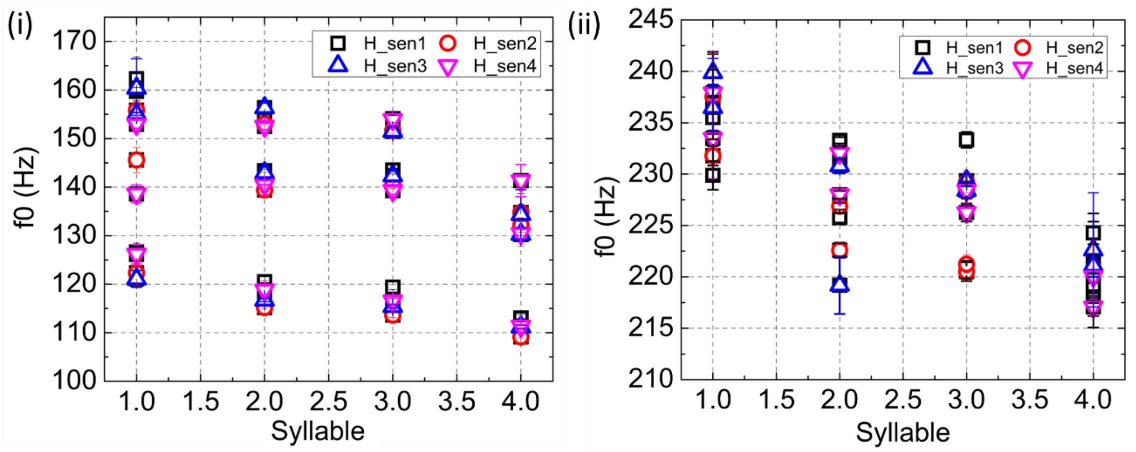

Figure 3.8: Time-normalized average f0 (in Hertz) for different sentences containing H tone sequences (H_sen1 to H_sen4) for male (i) and female (ii) speakers. These sentences vary in terms of number of syllables they contain. The x-axis represents the Syllable (S_k, where k= 1, 2, 3, and 4). Error bars depict standard errors aggregated over all the repetitions produced by all the speakers.

In simpler terms, the statistical modeling adopted in this study indicates that the linear models effectively depict the patterns in the H and L tone sequences. However, in the case of M tone sequences, a non-linear trend becomes apparent. Hence, the mathematical modeling is adopted to understand and identify the specific nature of this non-linearity and devise a method to assess it. The advantage of employing mathematical models over statistical ones is their ability to provide a more in-depth understanding of non-linear patterns. It is important to note that selecting the best mathematical model or fitting equation depends on the nature of the curves, the theoretical background, and the research context of this study.

Figure 3.5-3.7 (i-ii) exhibits the f0 trend across the normalized time for both male and female speakers. The syllable (S_k) is assigned to this time such that $S_k \in$ [1 + 10 (k-1), 10 k], for k = 1, 2, 3, 4, 5, and 6, denoting the Syllable no. **Figure 3.8(i-ii)** displays the average f0 for sentences containing H tone sequences as a function of S_k, where k ∈ [1, 4]. The length of these sentences varies in terms of the number of syllables they contain. The f0 decreasing trend is consistent across the male and the female speaker data. The f0 values fluctuate between 215-245 Hz for females and 110-170 Hz for males. Similar results are observed for the L tone sequences when the f0 is plotted as a function of S_k, with k ∈ [1, 5]. The f0 values vary within the range of 168-228 Hz for females and 90-126 Hz for males [see **Figure 3.9(i-ii)**].

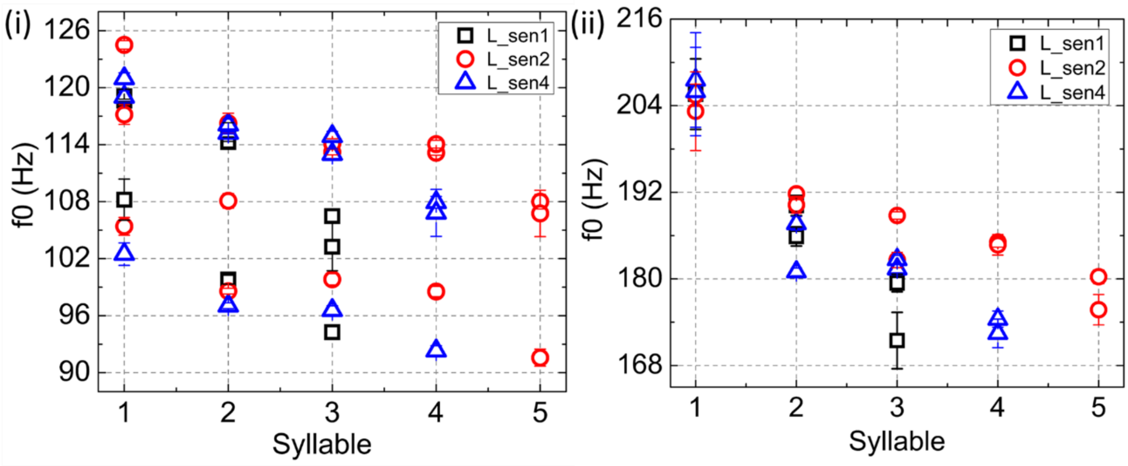

Figure 3.9: Time-normalized average f0 (in Hertz) for different sentences containing L tone sequences (L_sen1, L_sen2, and L_sen4) for male (i) and female (ii) speakers. These sentences vary in terms of number of syllables they contain. The x-axis represents the Syllable (S_k, where k = 1, 2, 3, 4, and 5). Error bars depict standard errors aggregated over all the repetitions produced by all the speakers

The visual observation of the f0(S_k) depicted in **Figures 3.8-3.9(i-ii)** reveals two distinct fitting equations: linear and exponential decay. The linear equation includes f0(S_k) = $\beta_1 S_k + \beta_0$, where f0(S_k) represents the dependent variable f0 in the y-axis, S_k is the independent variable in the x-axis, and $\beta_1$and β_0 are the coefficients akin to the slope and y-intercept, respectively. The exponential decay equation includes f0(S_k) = A exp(-S_k/B) + A_0, where A, B, and A_0 are the coefficients. A is the vertical scale of the decay, representing the maximum value that f0(S_k) can reach; B is the decay constant influencing the rate at which the function occurs. f0(S_k) decreases with increasing S_k. The larger value of B leads to a slower decay. Finally, A_0 is the offset representing any constant vertical shift to the entire function. The negative sign in the equation represents the decreasing trend of f0(S_k).

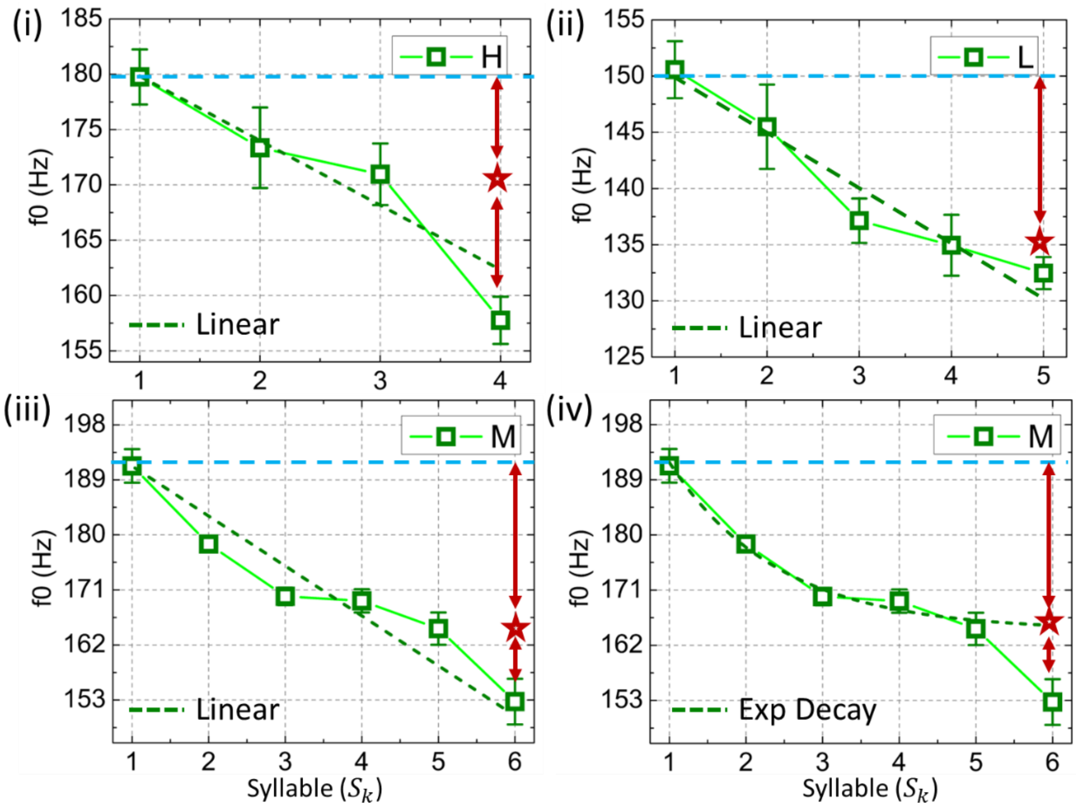

Figure 3.10: Time-normalized average f0 for all sentences featuring H, L, and M tone sequences produced by all the subjects is displayed in (i-iv) as a function of syllables. The corresponding green dashed lines represent the linear and the exponential decay fitted lines. The blue dashed line represents the highest f0. The red lines signify the vertical distance between the reference (calculated from the fit and denoted with a red star) and the actual final point.

Figure 3.10 (i-iv) shows the averaged f0(**S_k**) for all sentences featuring H, L, and M tone sequences uttered by all the subjects, denoting k ∈ [1, 4] for the H tone sequences, k ∈ [1, 5] for the L sequences, and k ∈ [1, 6] for the M sequences, respectively. The linear fit is feasible, whereas no convergence is achievable for the exponential decay fit in the case of the H tone sequences [see **Figure 3.10 (i)**]. The blue horizontal line represents the highest f0. This limitation arises not only from the narrow f0 range, which spans from 160 to 180 Hz, i.e., ~20 Hz, but also from the constraint on the syllable index (**S_k**), which is limited to 4. Conversely, both equations successfully fit the data for the M tone sequences. The f0 variation is similar for the L sequences, i.e., ~22 Hz for k ∈ [1, 5] [see **Figure 3.10(ii)**], whereas f0 variation increases to ~40 Hz with the k = 6 for M tone sequences [see **Figure 3.10 (iii-iv)].**

This analysis examines the declination slopes and their nature within each tone sequence. To achieve this, each tone sequence is individually fitted in both linear and non-linear (exponentially decay) equations. Notably, the fitting is performed for f0($\mathbf{S_k}$ - 1), excluding the last (final) point at ($\mathbf{S_k}$) to compare it with the reference point later. The reference point, denoted by a red star in Figure 7(i-iv), is calculated using the fitting parameters and the respective fitting equation. For instance, in the H tone sequences, the parameters for the linear fit are $\boldsymbol{\beta_1}$= -4.44 ± 0.91 Hz and $\boldsymbol{\beta_0}$= 183.83 ± 1.92 Hz, with an adjusted R-square value of 0.91, indicating a well-fitted curve. The reference point is calculated from f0($\mathbf{S_k}$= 4) = -(4.44 × 4) + 183.83 to assess whether the final point contributes to a lowering effect, resulting in approximately 166 Hz. The deviation is approximately 6 Hz from the actual final point (~160 Hz). This minimal difference suggests no pronounced final lowering effect in this context.

A similar analysis has been conducted for the other cases, including all L and M tone sequences shown in **Figure 3.10(ii-iv).** For sentences with all L tone sequences, the linear fit yields parameters $\boldsymbol{\beta_1}$ = -5.77 ± 1.23 Hz and $\boldsymbol{\beta_0}$ = 155.60 ± 3.47 Hz, with an adjusted R-square of 0.88, suggesting a relatively weaker fit compared to the H tone sequences [see **Figure 3.10(ii)**]. These negative values further validate the existence of declination in Chokri, as indicated in the H and L tone sequences [see **Figure 3.10(i-ii)**].

The linear fit equations of f0 for sentences featuring all M tone sequences give rise to lower R-squared values (~0.465) [see **Figure 3.10(iii)**]. The R-squared values increase significantly when an exponential decay equation is employed. The improved R-squared value becomes 0.85. The non-linear fitting is also performed for different orders of polynomial equations f0($\mathbf{S_k}$) = $\mathbf{a_n S_k^n}$+ $\mathbf{a_{n-1} S_k^{n-1}}$+ ... +$\mathbf{a_2 S_k^2}$+ $\mathbf{a_1 S_k}$+ $\mathbf{a_0}$, where $\mathbf{a_n}$, $\mathbf{a_{n-1}}$, ..., $\mathbf{a_2}$, $\mathbf{a_1}$, $\mathbf{a_0}$ are the coefficients of the polynomial. $\mathbf{S_k}$ is the variable, and n is the polynomial degree, indicating the highest power of $\mathbf{S_k}$ in the equation. It is observed that the R-squared value increases from 0.90 to 0.95 for n = 2 and n = 3, respectively. However, this value is not completely reliable. In evaluating the exponential versus polynomial fit in this context, it is crucial to understand that the exponential fit quantities naturally decrease over $\mathbf{S_k}$, aligning with our data's expected behavior. Additionally, B, the decay constant, affects the rate at which f0($\mathbf{S_k}$) decreases with increasing $\mathbf{S_k}$. The larger value of B leads to a slower decay. For the individual M tone sequences (M_sen1 and M_sen2) and the averaged M tone sequences, B is 0.75 ± 0.45, 0.87 ± 0.22, and 0.80 ± 0.32, respectively. The R-squared values are 0.81 for M_sen1 and 0.92 for M_sen2, which becomes 0.85 for the average M tone sequences. On the other hand, the polynomial fit provides flexibility with the increased R-squared values; however, it is risky since it may incorporate overfitting with high-degree polynomials, and there is the unreliability of extrapolation beyond the $\mathbf{S_k}$ range, which is important in this study. An influence of earlier data points diminishing over $\mathbf{S_k}$ visually aligns with the expected behavior of the f0 in sentences featuring all M tone sequences [see **Figures 3.5-3.6(i-ii)**]. Despite having a slightly lower R-squared value of 0.85, the exponential decay model demonstrates a better conceptual fit for f0($\mathbf{S_k}$).

Extrapolating the exponential curve to the reference point reveals that the final point is relatively lower than the fitted point (approximately 15 Hz), implying a final lowering in the syllable sequences featuring all M tones in Chokri, as illustrated in Figure 3.10(iii-iv). Therefore, the linear fit equation works for both H and L tone sequences, whereas the exponential decay fit equation is valid for the M tone sequence only. Since the k in $\mathbf{S_k}$ varies for each tonal sequence by one, it is hard to generalize its effects. Notably, the final lowering is only possible in the M tone sequences. The declination observed in Chokri does not depend on particular tonal sequences (H, M, and L).

3.3.5. Downstep in Chokri

This study further examines the effect of an intervening L tone in the H tone sequences to explore the potential occurrence of automatic downsteps in Chokri. This exploration is achieved through the sentences containing different combinations of syllables- HLH, HHLH, HLHH, HHLHH. Observing the time normalized f0 contours suggests that the intervening L can influence the preceding (H_p and H_2p) and the following syllable (H f and H 2f) containing an H tone. The legends, H_2p and H_2f, represent the location of a second preceding and following syllable featuring an H tone from the position of an intervening L tone (and not necessarily the immediate preceding syllable containing an H tone), respectively. In **Figure 3.11(i)**, the time-normalized average f0 is presented, averaging across all repetitions produced by all the subjects for the sentences containing the syllable sequence of HLHH, HHLH, and HHLHH, respectively.

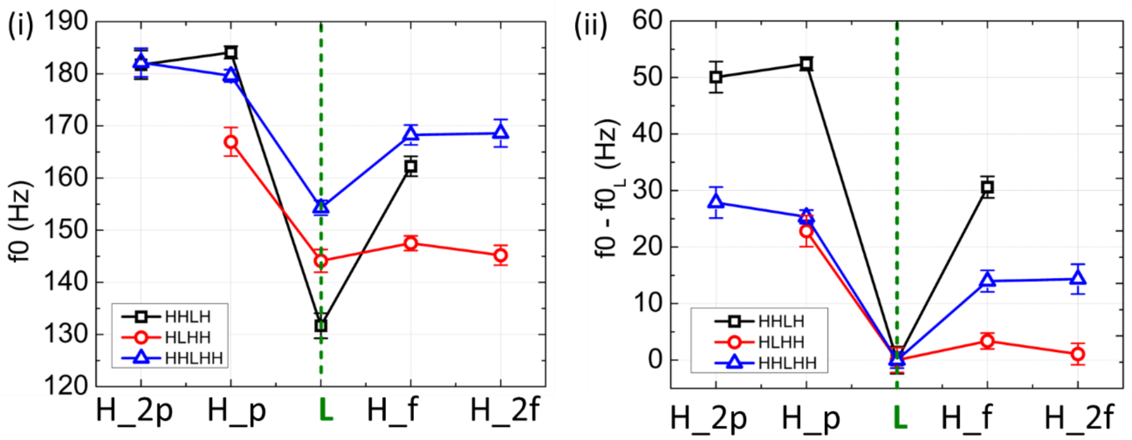

Figure 3.11: Time-normalized average f0 contours showing the effect of an intervening L between consecutive H tone in the syllable sequences featuring HHLH, HLHH, and HHLHH: (i) depict the time normalized f0 contours, and (ii) showcases the transformed f0 contours ($f0_L$). The transformed value is obtained by subtracting the f0 of the intervening L from each f0. The green dashed line is a reference point, distinguishing the preceding (H_p and H_2p) and following (H_f and H_2f) tone syllables.

The raw f0 values depicted in **Figure 3.11(i)** confirm that the intervening L tone predominantly affects the f0 of the immediately preceding and following H tone syllable (H_p and H_f) in the syllable sequence of HLHH, HHLH, and HHLHH, respectively. A reduction of f0 values in the immediate preceding H tone syllable of an intervening L tone syllable (depicted as H_p) is attributed to a possible declination effect. However, the second preceding H syllables- H_2p and H_2f, drawn in the sentences containing the syllable sequences of HHLH and HHLHH, are the least affected. Notably, a stiff fall in pitch occurs at the intervening L syllable, followed by a linear f0 contour. This phenomenon is distinctly observable in the HLHH and HHLH tonal sequences, confirming an automatic downstep in Chokri.

However, the raw f0 alone does not reveal the impact of an intervening L tone on these syllable sequences. The f0 is mathematically transformed to quantify this impact. This transformation ($f0_L$) is obtained by subtracting the f0 of the intervening L (~130 Hz, ~145 Hz, and ~155 Hz) from all the f0 ranges for HHLH, HLHH, and HHLHH, respectively. The y-axis of 8(ii), marked as f0 - $f0_L$, depicts the calculated values and indicates the impact of the intervening L tone on the neighboring H tone syllables in each sentence. The hierarchy of this intervening L tone's impact on the following H tone syllable is observed to be in the following order– HHLH (the impact is approximately 30 Hz) > HHLHH (the impact is approximately 15 Hz) > HLHH (the impact is approximately 5 Hz). On the other hand, the hierarchical order of the intervening L tone's impact on the preceding H tone syllable is observed to be in the following order– HHLH (the impact is approximately 50 Hz) > HHLHH HLHH (the impact is approximately 30 Hz for each syllable type). This trend approves the presence of automatic downstep borne out due to the influence of an intervening L tone.

3.4. Discussion

This study effectively addresses all three research questions: i) Given the presence of phrasal tones constituting APs in Sylheti, how is the surface realization of these tones influenced by the f0 downtrend that spans over IPs? Do lexical tones impact the downstepping of f0 peaks of phrasal tones? The scalings of f0 peaks created by the Ha AP boundary tones are subject to global downward movement of the f0 contour. The initial peak manifests with the highest f0 value, and each consecutive peak is realized with reduced f0 values than the previous one. The effect of lexical tones of IP medial roots is limited to the scaling of L* and the interpolating contour joining the L* and Ha. They do not change the specification of the phrasal tones, which means that downstep properties of f0 peaks in the language are not affected by lexical tones. (ii) Given Chokri's intricate nature of lexical tones (five-way tonal contrasts), to what extent are these tones affected by post-lexical downtrends of f0? The experimental results discussed above confirm that the post-lexical intonational effects like declination and automatic downstep are present in this language. However, the resulting gradual reduction of f0 is not present to the extent that would obliterate the lexical tones. iii) What types of downtrends manifest in Sylheti and Chokri? How do we quantitatively evaluate the presence of the downtrend effects in the language? The findings confirm that Chokri exhibits both declination and automatic downstep, while final lowering is restricted to all M tone sequences. Sylheti, on the other hand, has f0 peak downstep and final lowering. The integration of different mathematical fitting equations confirms these findings. Finally, (iv) What is the precise nature of downward f0 slopes resulting from downtrend effects? And how can these be modeled effectively? The distribution of f0 contours in Chokri affirmed that the declination manifests as an initial steep fall in f0, particularly observable in the first syllable, followed by a gradual slowing of the adjacent syllables. A linear model effectively fits the trend for the sentences with all H and L tone sequences. On the other hand, an exponential decay model establishes the declination nature of the sentences featuring all M tone sequences. The application of linear effects for H and L tone sequences and non-linear effects for the M tone sequence is substantiated through statistical and mathematical modeling. Consequently, this study elucidates how utterance-level pitch influences the realization of lexical pitch. For Sylheti, the successful implementation of the Liberman and Pierrehumberts (1984, 1988) downstep ratio (r) to predict f0 values of peak 2 and lowering constant (l) to predict the final peak in utterances with more than 2 peaks confirm the phonological nature of downstep. It shows that the precise nature of f0 decay between peaks involves a phonologically specified rate of change.

The outcome of this work seeks to fill in the need for quantitative modeling for enhanced verification of downtrend properties in natural languages. Devising apt mathematical models that best describe the global f0 slopes has complemented our initial postulations about the downtrend types in Sylheti and Chokri by establishing the precise nature of the f0 contours. Such an approach is also instrumental in bringing intonational phonology closer to speech technology, as our findings are expected to facilitate better systems for intonational aspects of speech synthesis and recognition. Most works on downtrends primarily explore the phonetic and phonological properties of various downtrend types. They are based on the measurement of f0 on syllables at different positions across an utterance as observed in Mambila (Connell, 2001), Igbo (Laniran, 1990), Cantonese (Johnson, 1986), Tswana (Zerbian and Kügler, 2015), to name a few. However, drawing mathematical modeling to establish the nature and types of the downtrend is still relatively scarce. Liberman and Pierrehumbert in 1984 showed that downdrift in English is realized as a decaying exponential, wherein the f0 slope becomes progressively less steep towards the later part of the phrase. In 2018, Myers (Scott et al., 1999) demonstrated that the downtrend of f0 in Chichewa could be evaluated using additive linear modeling and did not require exponential decay. However, the analysis is performed from a statistical modeling perspective, not the actual mathematical fitting. Furthermore, in other languages, such as Mandarin, it is exhibited that the declining f0 of sequences of tone1 (H) is steepest at the start of the utterance and is modeled with an exponential decay (Shih, 1998). However, this argument is primarily based on least square modeling, where the correlation is obtained in the utterances based on the correlation between the proceeding and following tone sequences.

The approach adopted in this study shares foundational concepts with those presented in Liberman and Pierrehumbert (1984). Quantitative and/or mathematical modeling of f0 features enables straightforward computation and manipulation of pitch contours for purposes of analysis and speech technology development. Offering a compact representation of the intonation contour makes them efficient for storage and computation. The systematic evaluation of fitting equations for all H, L, and M tone sequences in Chokri by comparing the linear and exponential decay modeling is a novel attempt made in this analysis. It is to be noted that a lower final point in the f0 vs. syllable or time plot does not necessarily provide evidence of the final lowering. Instead, this determination relies on factors such as total time, syllable count, and the nature of the fitted mathematical equation. The convergence of exponential decay may vary across data, and linear fits may not universally perform well. Judgment calls depend on domain knowledge and context.

While this study offers valuable insights into the independent dynamics of downtrends, further research exploring their interaction with other intonational functions of f0, such as focus and phrasing, will contribute to a more comprehensive understanding of this phenomenon.

References

Appels, R. (1985). Een productiemodel voor declinatie en onderzoek naar declinatie-resets

binnen spraakuitingen. Internal Report no. 498, Institute for Perception Research, Eindhoven.

Arvaniti, A. (2004) Final lowering: Fact, artifact or dialectal variation?. *J. Acoust. Soc. Am.* 1 October 2004; 116 (4_Supplement): 2644. https://doi.org/10.1121/1.4785546

Beckman, M. E. & J. Pierrehumbert. (1986). Intonational structure in Japanese and

English. Phonology Yearbook 3: 255-310.

Becker, L.A. (1979). 'Once more on the nature of downdrift.' *Studies in African Linguistics*, Vol 10. 3, 233–245.

Boersma. P., Weenik D (2012). Praat: doing phonetics by computer (Version 5.3.04_win64) [Computer program].

Bolinger, D. (1978). Intonation across languages. In J. Greenberg (ed.), *Universals of Human Language*, Volume 2 (*Phonology*), pp. 471–524. Stanford: Stanford University Press.

Connell, B. (2001). Downdrift, Downstep, and Declination. Typology of African Prosodic Systems Workshop Bielefeld University, Germany

Connell, B. (2017). Tone and Intonation in Mambila. In L. J. Downing, & A. Rialland (Eds) *Intonation in African Tone Languages*(pp). De Gruyter Mouton

Connell, B. (2006). Four tones and downtrend: a preliminary report on pitch realization in Mambila, a language with four level tones, in *27th ACAL*, University of

Florida, Gainesville, 29-31 Mar 1996 and *the LAGB*, University of Sussex, Brighton, 11-13 Apr 1996.

Connell, B. & Ladd R. (1990). Aspects of pitch realization in Yoruba. *Phonology* 7: 1-30

Cohen, A., & ' t Hart, J. (1967). On the anatomy of intonation. *Lingua, 19* 177–192.

Cooper, W. E. and J. M. Sorensen. (1981). *Fundamental Frequency in Sentence Production*. Springer, Berlin.

Deguchi, M., & Kitagawa Y. (2002). Prosody and Wh-Questions. In M. Hirotani (Ed.), Proceedings of the North East Linguistics Society (pp. 73-92).

Estebas-Vilaplana, E. (2000). Peak F0 downtrends in Central Catalan neutral declaratives. *Speech, Hearing, and Language: Work in Progress, 16*, 41. London

Flynn, C. Y. (2003). Intonation in Cantonese. LINCOM Studies in Asian Linguistics 49

Fox, A., Luke, K., & Nancarrow, O. (2008). Aspects of intonation in Cantonese. *Journal of Chinese Linguistics, 36*(2), 321-367.

Gussenhoven, C. and A. Rietveld (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics* 16, 355–369.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hieber, D. W. (2016). The cohesive function of prosody in Ékegusií (Kisii) narratives: A functional-typological approach. *UC Santa Barbara*. ProQuest ID: Hieber_ucsb_0035N_12906. Merritt ID: ark:/13030/m510223m. Retrieved from https://escholarship.org/uc/item/4wp128k9

Hyman, Larry M. 2001. Tone systems. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), Language typology and language universals: An international handbook, 1367–1380. Berlin: Walter de Gruyter.

Ishihara, S. (2007). Major phrase, focus intonation, multiple spell-out (MaP, FI, MSO). *Linguistic Rev.* 24, 137–167. doi: 10.1515/TLR.2007.006

Johnson K. (1986) Tone and intonation in Cantonese. *J. Acoust. Soc. Am.* 1 May 1986; 79 (S1): S37. https://doi.org/10.1121/1.2023194

Kawahara, S., & Shinya, T. (2008). The intonation of gapping and coordination in Japanese: Evidence for international phrase and utterance. Phonetica, 65, 62-105. doi:10.1159/000130016

Kula, N. & Hamann, S. (2017). Intonation in Bemba. In L. Downing & A. Rialland (Ed.), *Intonation in African Tone Languages* (pp. 321-364). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110503524-010

Ladd, R. (1984). Declination: A review and some hypotheses. *Phonology Yearbook*, 1:53–74

Laniran, Yetunde (1992). Phonetic aspects of tone realization in Igbo. Progress Reports from Oxford Phonetics5. 35-51.

Laniran, Y. d. (1990). Intonation in tone languages: The Yoruba example. Ph.D. thesis, Cornell University.

Lieberman, P. (1967). *Intonation, perception, and language: Cambridge*, MA: MIT Press.

Lieberman, P. (1980). The innate, central aspect of intonation. See Waugh and Van Schooneveld (1980), pp. 187–199.

Liberman M &. Pierrehumbert J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oehrle, editors, *Language SoundStructure*, pages 157–233. M.I.T. Press, Cambridge, Massachusetts.

Maeda, S. (1976). *A Characterization of American English Intonation*. PhD thesis, MIT.

Myers, S., Selkirk, E. & Fainleib, Y., (2018) "Phonetic implementation of high-tone spans in Luganda," *Laboratory Phonology* 9(1): 19. doi: https://doi.org/10.5334/labphon.101

Ohala, J. (1981). The listener as a source of sound change. In *Papers from the Parasession on Language and Behavior* (pp. 178–204). Chicago Linguistic Society.

Ohala, J., & Ewan, W. G. (1973). Speed of pitch change. *Journal of the Acoustical Society of America*, *53*(1), 345. https://doi.org/10.1121/1.1912878

Pierrehumbert, J. (1979). The Perception of Fundamental Frequency Declination. *Journal of the Acoustical Society of America*, Vol. 66, No. 2, pp. 363–369.

Pierrehumbert, J & Beckman, M. (1988). *Japanese Tone Structure*. The MIT Press, Cambridge, Massachusetts.

Prieto, P., C. Shih, and H. Nibert. 1996. Pitch downtrend in Spanish. Journal of Phonetics 24: 445-473.

Rialland, A. (1997). Le parcours du "Downstep" ou l'evolution d'une notion. *Bulletin de la Société de Linguistique de Paris* 92:207–243.

Shih, C. (1998). Declination in Mandarin. *Intonation: Theory, Models and Applications: Proceedings of an ESCA Workshop,* September 18 - 20, Athens, Greece

Snider, K. L. (1998). Phonetic Realisation of Downstep in Bimoba. *Phonology*, *15*(1), 77–101.

Stewart, J.M. (1983). 'Downstep and Floating Low Tones in Adioukrou.' *Journal of African Languages and Linguistics*, Vol 5,57-78.

Strik, H. and L. Boves (1995). Downtrend in f0 and P*sb Journal of Phonetics* 23, 203–220.

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

Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *The Journal of the Acoustical Society of America* 83, 1536–1552.

Truckenbrodt, H. (2004). Final lowering in non-final position. Journal of Phonetics. 32. 313-348. 10.1016/j.wocn.2003.11.001.

Wang, Bei & Kügler, Frank & Genzel, Susanne. (2022). The interaction of focus and phrasing with downstep and post-low-bouncing in Mandarin Chinese. *Frontiers in Psychology*. 13. 10.3389/fpsyg.2022.884102.

Welmers, W. (1973). *African Language Structures*. Berkeley: University of California Press.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *J. Phon*. 27, 55–105. doi: 10.1006/jpho.1999.0086

Xu, Yi. (2013). ProsodyPro - A tool for large-scale systematic prosody analysis. Proceedings of the TRASP Conference. 7-10.

Zerbian, S., & Kügler, F. (2015). “Downstep in Tswana (Southern Bantu).” in *Paper Presented at the ICPhS*, Glasgow.

Zerbian, S., & Kügler, F. (2021). Sequences of high tones across word boundaries in Tswana. *J. Int. Phon. Assoc.* 1–22. doi: 10.1017/S0025100321000141