Computational Approach to Bengali Stress

Shohini Bhattasali*

1 Introduction

In this work, my goal is to train a computational model to detect stress in Bengali using data from a speech corpus and then compare my results against existing accounts of Bengali stress which differ in their analyses.

Stress refers to the relative prominence of portions of an utterance (Liberman and Prince 1977). It has also been defined as the linguistic manifestation of rhythmic structure (Liberman 1975, Liberman and Prince 1977). Hayes (1995) explains this further and states that in stress languages, "every utterance has a rhythmic structure that serves as a framework for that utterance's phonological and phonetic realization" (8). However, any formal theory of stress has to account for considerable cross-linguistic variation and the different acoustic correlates of stress such as duration and intensity.

Hayes (1980) proposed that stress patterns can be classified into two different types: quantity-insensitive stress systems and quantity-sensitive stress systems. Quantity-insensitive systems are those in which syllable weight is not relevant in conditioning stress placement. Gordon (2011) gives the example of the Australian language Maranungku as an example of a quantity-insensitive stress system. In this language, the primary stress falls on the first syllable of a word and secondary stress docks on the remaining odd-numbered syllables, as seen in examples (1 – 4) from Tryon 1970.

(1) ˈtiralk
'saliva'

(2) ˈmæræˌpæt
'beard'

(3) ˈjaŋarˌmata
'the Pleiades'

(4) ˈŋaltiˌritiˌti
'tongue'

Conversely, in quantity-sensitive stress systems, stress is sensitive to syllable weight. Yana (Sapir and Swadesh 1960) is an example of one such language. In this language, stress falls on the leftmost heavy syllable (CVV or CVC), otherwise the initial syllable receives stress, as seen in examples (5) – (8).

(5) ˈp'udiwi
'women'

(6) siˈbumk'ai
'sandstone'

(7) suˈk'oːniyaː
'name of Indian tribe'

*I would like to thank Mats Rooth, Draga Zec and Sam Tilsen for all their guidance and feedback.

(8) tsiniˈjaː
‘no’

In this work I am focusing on Bengali, which has been classified as both a quantity-insensitive system by Hayes and Lahiri (1991) and as a quantity-sensitive system by Shaw (1984). There is no agreement on the stress pattern in Bengali but all studies agree that stress in Bengali is predictable (Hayes and Lahiri 1991, Shaw 1984, Das 2001).

The paper is organized as follows: in Section 2, I provide a brief overview about Bengali and the existing accounts of stress in Bengali and in Section 3 I explain the main objective behind this study. Next, I discuss relevant background information about my approach, the speech corpus, toolkits and give an overview of how my model is trained to detect stress cues in Bengali in Section 4. Lastly, I present my results in Section 5 and end with a discussion of the results in Section 6.

2 Background

2.1 Bengali

As per Ethnologue (Lewis et al. 2015), Bengali is an Indo-Aryan language with over 180,000,000 speakers, spoken mainly in India and Bangladesh.

Figures 1 and 2 illustrate the phonemic inventory for Bengali (adapted from Khan 2011). The consonant inventory consists of stops in three series (voiceless, voiced, aspirated) and three places of articulation (labial, coronal, dorsal). There are seven vowels and the vowel quality does not change in stressed positions.

<table>
<tr><th></th><th>Bilabial</th><th>Labio-dental</th><th>Dental</th><th>Alveolar</th><th>Post-alveolar</th><th>Velar</th><th>Glottal</th></tr>
<tr><td>Plosive</td><td>p b
pʰ bʰ</td><td></td><td>t̪ d̪
t̪ʰ d̪ʰ</td><td></td><td>ʈ ɖ
ʈʰ ɖʰ</td><td>k g
kʰ gʰ</td><td></td></tr>
<tr><td>Affricate</td><td></td><td></td><td></td><td></td><td>ʧ ʤ
ʧʰ ʤʰ</td><td></td><td></td></tr>
<tr><td>Nasal</td><td>m</td><td></td><td colspan="3">n</td><td>ŋ</td><td></td></tr>
<tr><td>Fricative</td><td></td><td></td><td colspan="2">s</td><td>ʃ</td><td></td><td>h</td></tr>
<tr><td>Approximant</td><td></td><td></td><td colspan="3">r</td><td></td><td></td></tr>
<tr><td>Lateral approximant</td><td></td><td></td><td colspan="3">l</td><td></td><td></td></tr>
</table>

Figure 1: Bengali consonant inventory.

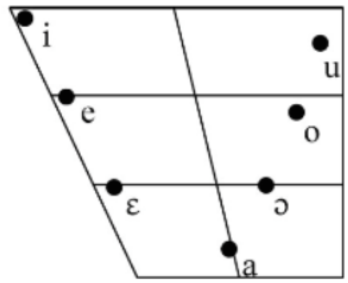

Figure 2: Bengali vowel inventory (Khan 2011).

According to Grierson's (1930) survey, there are 8 dialects of Bengali, including some (such as the ones spoken in Chakma, Chittagonia, Sylheti) which have been considered different languages with differing phonemic inventories, allophony, and inflectional morphology (Gordon 2005). In this paper, I only use data from the Standard Colloquial Bengali (SCB) dialect of Bengali.

Bengali can be divided intro three lexical strata, and syllabification patterns vary across the different strata in Bengali. Modelled after Ito and Mester's (1995) core-periphery analysis of the lexical strata in Japanese, Kar (2009) presents a stratification system for the Bengali lexicon. In his analysis, Kar explains how the syllable structure constraints are ranked differently at each stratum to give us different syllable shapes for each lexical stratum. He proposes that Bengali can be divided into the following three strata:

- *Tadbhaba*: These are the native Bengali words, rooted in Sanskrit and Prakit and this stratum is named the Native Bangla (NB) stratum. E.g., *kaʈ*^{h} 'wood', *p*^{h}*ul* 'flower', etc. The syllable shape is of the form CV(C).

- *Tatsama*: These are words borrowed directly from Sanskrit and this stratum is named the Sanskrit Borrowings (SB) stratus. E.g., *gram* 'village', *kobi* 'poet', *ʧahid̪a* 'demand', *snan* 'bath' etc. The syllable shape is of the form C(C)V(C).

- *Deshi o Bideshi*: These are words borrowed from Indian (*deshi*) and foreign (*bideshi*) languages and this stratum is named the Other Borrowings (OB) stratum. E.g., *anarɔʃ* 'pineapple' (from Portugese *ananas*), *burʤoa* 'middle class' (from French *bourgeois*), *ast̪e* 'slowly' (from Persian *ahistah*), *rikʃa* 'rickshaw' (from Japanese *rikiʃa*), *haʃpat̪al* 'hospital' (from English), *hɔrot̪al* 'strike' (from Gujarati *hɔt̪tal*). The syllable shape is of the form C(C)V(C)(C).

Thus, we can see that in NB stratum, only syllables with simple onsets and simple codas are allowed; in the SB stratum, syllables with complex onsets and simple codas are allowed; and in the OB stratum, the syllable with complex onsets and complex codas are allowed. Chatterji (1921) mentions that distribution of stress varies between native Bengali words and Sanskrit borrowings.

2.2 Word-level Stress in Bengali

While there is general consensus that Bengali has fixed and non-contrastive word-level stress, more than one analysis has been proposed in the literature. There are at least two formal accounts about the pattern of primary stress in Bengali and each of them paints a different picture.[1] They differ in their approaches (namely, a quantity insensitive account versus a quantity sensitive account of stress) and also regarding the default placement of stress.

Proposal 1: Hayes and Lahiri (1991) follow a quantity insensitive account of stress, stating that the first syllable is stressed in Bengali and it is an inviolable rule. Thus, in examples (9) – (14), the first syllables *ba*, *ʃɔ*, *a*, *go*, *ɖʰu*, and *ma*, respectively receive stress. They base their analysis on previous accounts of Chatterji (1921) and Klaiman (1987). Other sources agree with their explanation (Anderson 1920, Goswami 1944, Ferguson and Chowdhury 1960, Bykova 1981, Kawasaki and Shattuck-Hufnagel 1988, Fitzpatrick-Cole and Lahiri 1997, Lahiri and Fitzpatrick-Cole 1999, Michaels and Nelson 2004, Selkirk 2007).

(9) ˈbari
‘house’

(10) ˈʃɔmaj
‘society’

(11) ˈanɔnd̪o
‘happiness’

(12) ˈgolapi
‘pink’

(13) ˈɖʰukeʧʰilo
‘entered’

(14) ˈmakʰieʧʰilen
‘you had mashed’

Proposal 2: Shaw (1984) expresses a view that departs from the standard account, as seen described above. He presents a quantity sensitive account of stress in Bengali, proposing that stress falls only on the first syllable if it is heavy, otherwise it is assigned to the second syllable. In this account, a closed syllable (CVC) constitutes a heavy syllable. In (15) – (17), the first syllables are heavy (closed) and thus, *sɔŋ*, *bak*, and *d̪ap* receive primary stress respectively, whereas in (18) – (20) the first syllable is not heavy, so the second syllable *kaʃ*, *bi*, and *buʤ* receive stress, respectively.

(15) ˈsɔŋʃar
‘family’

(16) ˈbakʃoʈa
‘the box’

(17) ˈd̪apt̪ar
‘office’

[1] Secondary stress in Bengali is not discussed in this paper since there is debate regarding its existence in Bengali.

(18) aˈkaʃ
'sky'

(19) koˈbit̪a
'poem'

(20) ʃoˈbuʤ
'green'

If we compare Proposal 1 and Proposal 2, there is an overlap in their prediction for the distribution of stress. If the first syllable of a word is heavy, then both proposals predict that it receives primary stress, as in (21). However, when the first syllable is light, the two accounts diverge. According to Proposal 1, the first syllable would still receives stress, as in (22). On the other hand, according to Proposal 2, the second syllable would receive stress instead of the first syllable as in (23).

(21) ˈkalpɔnik
'imaginary'

(22) ˈbat̪aʃa
'type of candy'

(23) baˈt̪aʃa
'type of candy'

The two proposals can also be compared with respect to a metrical foot-based approach (Hayes 1995). Both of these proposals would use the constraint ranking ALL-FT-L >>PARSE-SYL. However, the foot forms would be different. For Proposal 1 the foot would be a syllabic trochee with the form $^{ˈ}(\sigma)\sigma\sigma\sigma$. On the other hand, for Proposal 2, the foot would be a quantity-sensitive iamb of the form $(\sigma^{ˈ}\sigma)\sigma\sigma$.

There could be various reasons for the discrepancy seen in these distribution patterns. Apart from Hayes and Lahiri (1991) who used duration and intensity as acoustic correlates of stress to analyze Bengali waveforms, the other studies were mostly impressionistic. They were based exclusively on auditory cues and did not consider phonetic correlates of stress to support their account. Additionally, all of the data does not come from the same dialect and regional variation could be influencing the proposed distribution.

3 Objective

Bengali is classified as a stress language (Hayes 1995) and stress can be word-initial or it can depend on the weight, i.e., heavy syllables receive primary stress (as described in Section 2.2). My goal is to use data from a speech corpus to provide evidence in support of one account over another using a computational model. Thus, I want to do the following:

- Train an ASR (Automatic Speech Recognition) system to detect stress cues in a speech corpus,
- Test the existing formal accounts of stress against empirical results from the speech corpus.

4 Methodology

The following subsections provide an overview about finite-state models, the Bengali speech corpus, toolkits, and the models used to detect stress cues.

4.1 Finite-state Approach

A finite state model makes use of the concept of finite state machines (FSM) or finite state automata (FSA). It is conceived as an abstract machine that can be in one of a finite number of states. In Figure 3 we have an example of a simple FSM with three states such that S_0 is the start state and S_2 is the final state. We could go directly from S_0 to S_2 or transition via S_1.

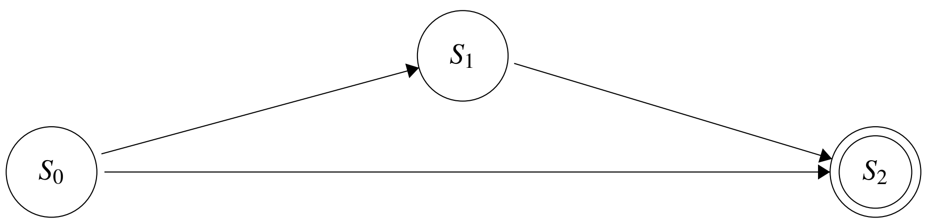

Figure 3: A FSM with 3 states.

Finite state models have been effectively used to represent various aspects of computational phonology and morphology. The application of finite-state methods in phonology and morphology was first proposed by Karttunen (1993), based on Johnson (1970) and Kaplan and Kay (1980)'s insight about phonological rewrite rules being represented as finite state operations. Karttunen gives us the following example of vowel back harmony in a FSM as seen in Figure 4. The upper and lower parts of the machine are represented as *x*:*y*, respectively and can correspond to the underlying form (*x*) and the surface representation respectively (*y*).

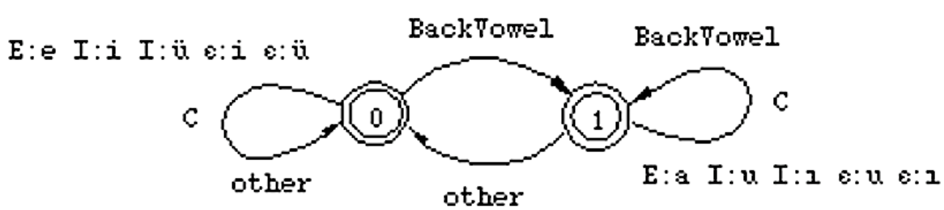

Figure 4: FSM for back harmony (Kartunnen 1993).

While there are no existing works incorporating a finite state approach to modelling stress, Karttunen's subsequent work on Finnish prosody utilized FSMs (Karttunen 2006) and Gildea and Jurafsky (1995)'s work on representing English phonological rules in a finite state model are some good examples of using finite state methods to give further insight into phonological patterning. For example, Figure 5 illustrates an FSM for *bat*, *batter*, and *band* with the underlying and surface representations.

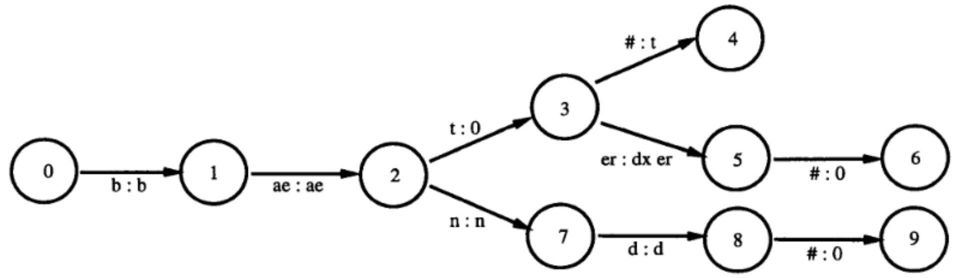

Figure 5: FSM for English words *bat*, *batter*, and *band* (Gildea and Jurafsky 1995).

4.2 Data

The data for this experiment comes from Shruti (Mandal et al. 2011, Das et al. 2011), a Bengali speech corpus. It has been recorded, transcribed, and maintained at Indian Institute of Technology, Kharagpur. The speech corpus consists of utterances from 34 native speakers of Standard Colloquial Bengali as they read out news articles from *Anandabazar Patrika*, a popular Bengali newspaper in Kolkata. It contains 7,383 unique sentences, a total of 22,012 words, and 21.64 hours of recorded speech. The percentages of male and female speakers are 75% and 25% respectively and their ages vary between 20 years to 40 years.

4.3 Toolkits

The two toolkits that I use for my experiment are OpenFST and Kaldi, discussed in the following subsections.

4.3.1 OpenFST

OpenFST is a weighted finite-state toolkit created by Allauzen et al. (2007). It consists of a library for constructing, combining, optimizing, and searching weighted finite-state transducers (FSTs). A finite state transducer is a finite state automaton whose state transitions are labeled with both input and output labels. A weighted transducer places weights on transitions in addition to the input and output labels. Mohri et al. (2002) explain that weights may include probabilities, duration, or any other quantity that accumulates along the paths to determine the overall weight of mapping an input sequence to an output sequence and thus, they explain that weighted finite state transducers are a natural choice to represent the finite state modelling prevalent in speech processing.

4.3.2 Kaldi

Kaldi, created by Povey et al. (2011), is a free, open source ASR toolkit for acoustic modelling research, written in C++, and it provides a speech recognition system based on finite-state transducers. It aims at converting speech signal into readable text in real time. To accomplish this goal, we give Kaldi the speech signal files, the transcripts of the corpora to be analyzed, along with a phonetic dictionary (a list of words with their phonetic transcriptions) for each corpus, and a list of phones of the language as input. The system will train the data, and will generate output files that contain the aligned utterances (wave files), as well as decoded files (i.e., text files) that contain the utterances decoded after having been aligned. Kaldi utilizes MFCCs (Mel Frequency Cepstral Coefficients) which are non-linearly-spaced frequency bands, which are a close approximation of the human auditory systems response (Hasan et al. 2004). It is compiled against the OpenFST library and uses Hidden Markov Models (HMMs) to train the data.

Finite state machines (discussed in Section 3.1) can be represented as HMMs and these HMMs are used to model systems that are assumed to be Markov Processes which contain hidden (unobserved) states. HMMs explicitly map between the acoustic correlates in a speech corpus and lexicons with stress markings, thereby providing a framework for modeling time-varying feature vectors of a speech sound. HMMs have been successfully used in speech recognition problems for a few decades now.

An HMM with a Gaussian Mixture Model (GMM) state-output distribution can be used to model the asymmetric and multi-modal data of the speech corpus. GMMs are a commonly used estimate of the probability density function used in statistical classification systems (Reynolds and Rose 1995). Thus, the HMM models the temporal data as a sequence of states and states are defined as separate GMMs and in this way, the HMM creates a sequence of GMMs to explain the input data, as seen in the figure below:

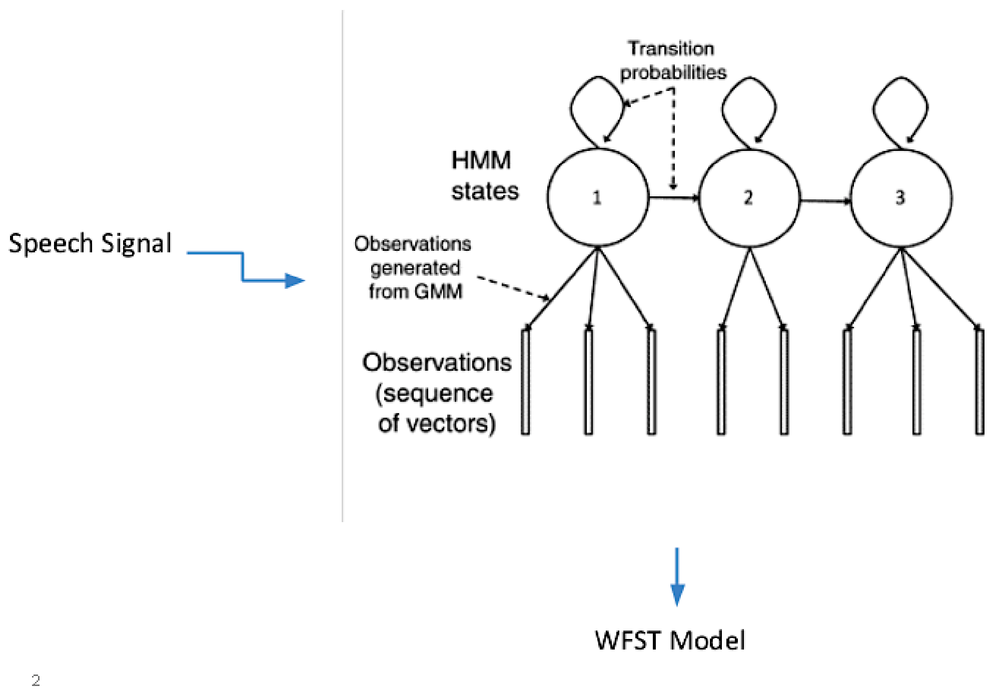

Figure 6: An HMM-GMM for speech recognition (Povey et al. 2011).

For the purposes of the present study, only the aligned files are relevant. Kaldi uses Baye's rule to predict the most likely utterance for a given speech signal (Povey et al. 2011):

$$P(S|audio) = \frac{p(audio|S)P(S)}{p(audio)}$$

where:

- P(S|audio) – the most likely sentence: given a test utterance, the S that maximizes P(S|audio) is picked;
- p(audio|S) – sentence-dependent statistical model of audio production, trained from data;
- S – the sequence of words;
- P(S) a language model (LM), i.e., an n-gram model or a probabilistic grammar;
- p(audio) – normalizer;

In Kaldi, strings are sequences of symbols: Kaldi uses Weighed Finite State Acceptors (WFSA) to accept strings, i.e., a string is accepted when there is a minimum-cost path containing the sequence of symbols from the initial to the final state of the machine. Kaldi's training and decoding algorithms use Weighted Finite State Transducers (WFST).

4.4 System Architecture

I train a HMM-GMM model to accurately align words from the speech corpus, Shruti to demonstrate the position of stress in Bengali. Using the ASR system Kaldi, the sound waves from the corpus are converted to their corresponding vector representations. Through OpenFST, a finite state model is trained on a phonetic dictionary to map these vector representations to the transcript of a utterance. The model outputs alignments to demonstrate the mappings.

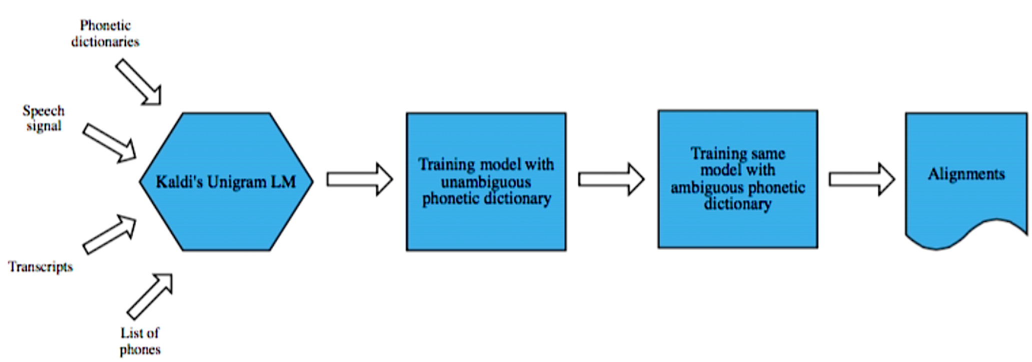

Figure 7: Overview of training process.

Extending this concept, I train two models iteratively on different phonetic dictionaries based on the Shruti speech corpus, as illustrated by the schematic diagram in Figure 7. Kaldi uses duration and intensity as the acoustic correlates of stress and stress is marked following the conventions of the CMU phonetic dictionary (Weide 1998) with ‘1’ beside a phoneme indicating primary stress.

Both the models are given an unambiguous phonetic dictionary at the first training stage. In stage one, Model 1 trains on a phonetic dictionary with the first syllable marked for stress and Model 2 trains on a phonetic dictionary with the second syllable marked for stress. Then, the trained models are given an ambiguous phonetic dictionary which has each word with its stressed variants and unstressed variant, as seen in Figure 8 and 9.

```
abadAna2 aa1 b oh dd A n
abadAna aa b oh dd A n
abadAna aa1 b oh dd A n
abadAna aa b oh1 dd A n
AbadAra2 A1 b dd A rr
AbadAra A b dd A rr
AbadAra A1 b dd A rr
AbadAra A b dd A1 rr
AbadArao2 A1 b dd A rr oh
AbadArao        A b dd A rr oh
AbadArao        A1 b dd A rr oh
AbadArao        A b dd A1 rr oh
AbadAre2 A1 b dd A rr ee
AbadAre A b dd A rr ee
AbadAre A1 b dd A rr ee
AbadAre A b dd A1 rr ee
Abaddha2 A1 b aa dd ddh oh
Abaddha A b aa dd ddh oh
Abaddha A1 b aa dd ddh oh
Abaddha A b aa1 dd ddh oh
abadhi2 aa1 b oh ddh i
abadhi  aa b oh ddh i
abadhi  aa1 b oh ddh i
abadhi  aa b oh1 ddh i
```

Figure 8: Phonetic dictionary for Model 1.

```
abadAna3        aa b oh1 dd A n
abadAna aa b oh dd A n
abadAna aa b oh1 dd A n
abadAna aa1 b oh dd A n
AbadAra3        A b dd A1 rr
AbadAra A b dd A rr
AbadAra A b dd A1 rr
AbadAra A1 b dd A rr
AbadArao3       A b dd A1 rr oh
AbadArao        A b dd A rr oh
AbadArao        A b dd A1 rr oh
AbadArao        A1 b dd A rr oh
AbadAre3        A b dd A1 rr ee
AbadAre A b dd A rr ee
AbadAre A b dd A1 rr ee
AbadAre A1 b dd A rr ee
Abaddha3        A b aa1 dd ddh oh
Abaddha A b aa dd ddh oh
Abaddha A b aa1 dd ddh oh
Abaddha A1 b aa dd ddh oh
abadhi3 aa b oh1 ddh i
abadhi  aa b oh ddh i
abadhi  aa b oh1 ddh i
abadhi  aa1 b oh ddh i
```

Figure 9: Phonetic dictionary for Model 2.

In stage two, Model 1 and Model 2 have to choose between the different versions based on what it has learned through the training data in the first stage to produce the alignments mapping the sound waves to the utterances. My hypothesis is that if the learning is entirely probabilistic, then Model 1 and Model 2 will consistently choose the variants of the word with the first syllable and the second syllable marked for stress respectively. However, if the models have learned to detect stress after the first round of training, it will predict a different distribution based on the data and thereby, it learns to predict the distribution of the stressed phoneme.

5 Results

The results from Kaldi are in the form of alignments which map the sound waves to the utterances. The images below show excerpts from the two sets of Kaldi alignments. The stress cues detected by Kaldi are highlighted in red.

```
shyamoshree-89 SIL A1_B chh_E rr_B A1_I tt_I rr_I ee_E tt_B u_I m
_I i_E n_B i1_I s_I ch_I i_I n_I tt_I ee_E b_B i_I sh_I rr_I A_I
m_E k_B aa_I rr_I oh_E SIL
shyamoshree-890 SIL bh_B aa_I dd_I rr_I oh_I l_I oh_I k_E k_B oh1
_I tth_I A_E tth_B ee_I k_I ee_E sh_B u_I n_I ee_I chh_I ee_I n_E
 SIL A1_B p_I n_I A_I rr_E s_B aa_I ^n_I g_I ee_E A1_B m_I A_I rr
_E kh_B u_I b_E SIL j_B A_I n_I A_I s_I oh_I n_I A_E SIL
shyamoshree-91 SIL tt_B A1_I rr_I p_I aa_I rr_E ee1_B i_E sh_B aa
_I h_I oh_I rr_I ee_E j_B i_I b_I i_I k_I A_I rr_E E1_B k_I T_I A
_E s_B aa_I n_I ddh_I A_I n_E k_B oh_I rr_I ee_E SIL n_B A_I oh_E
 SIL
shyamoshree-92 SIL A1_B s_I rr_I aa_I Y_I ee_I rr_E ch_B i1_I n_I
 tt_I A_E n_B i1_E SIL ch_B ou_I tt_I rr_I i_I sh_E n_B aa1_I m_I
 b_I oh_I rr_E gh_B aa_I rr_E tt_B oh_I m_I A_I k_I ee_E A_B m_I
i_E SIL ph_B ee_I rr_I oh_I tt_E p_B A_I i_I ee_E dd_B ee_I b_I o
h_E SIL
```

Figure 10: Alignments for Model 1.

```
shyamoshree-89 SIL A_B chh_E rr_B A_I tt_I rr_I ee1_E tt_B u_I m_
I i1_E n_B i_I s_I ch_I i_I n_I tt_I ee_E b_B i_I sh_I rr_I A1_I
m_E k_B aa_I rr_I oh1_E SIL
shyamoshree-890 SIL bh_B aa1_I dd_I rr_I oh_I l_I oh_I k_E k_B oh
_I tth_I A1_E tth_B ee_I k_I ee1_E sh_B u1_I n_I ee_I chh_I ee_I
n_E SIL A_B p_I n_I A1_I rr_E s_B aa1_I ^n_I g_I ee_E A_B m_I A1I
 rr_E kh_B u_I b_E SIL j_B A_I n_I A1_I s_I oh_I n_I A_E SIL
shyamoshree-91 SIL tt_B A_I rr_I p_I aa1_I rr_E ee_B i1_E sh_B aa
_I h_I oh1_I rr_I ee_E j_B i_I b_I i1_I k_I A_I rr_E E_B k_I T_I
A1_E s_B aa_I n_I ddh_I A1_I n_E SIL k_B oh_I rr_I ee_E SIL n_B 1
_I oh_E SIL
shyamoshree-92 SIL A_B s_I rr_I aa1_I Y_I ee_I rr_E ch_B i_I n_I
tt_I A1_E n_B i_E SIL ch_B ou_I tt_I rr_I i1_I sh_E n_B aa_I m_I
b_I oh1_I rr_E gh_B aa_I rr_E tt_B oh1_I m_I A_I k_I ee_E A1_B m_
I i_E SIL ph_B ee_I rr_I oh1_I tt_E p_B A1_I i_I ee_E dd_B ee_I b
_I oh_E SIL
```

Figure 11: Alignments for Model 2.

I compare both sets of alignments with respect to the percentage of stress cues detected. The percentages are given in the bar graph below:

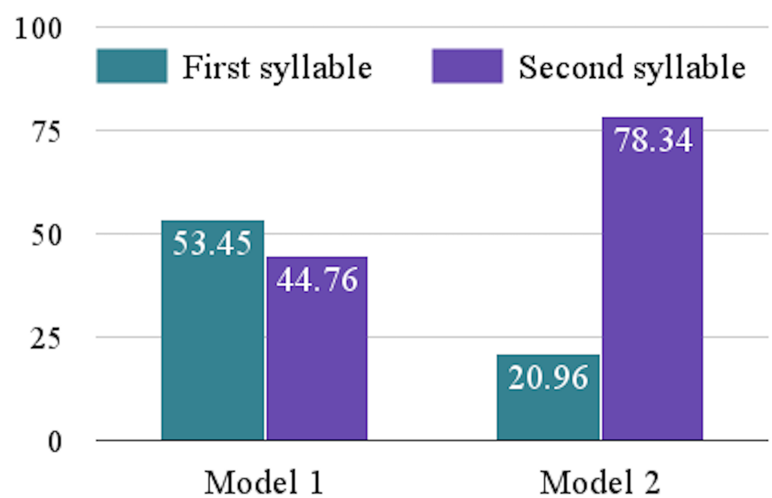

Figure 12: Comparison of stress placement between two models.

With the total of 7,934 utterances, Model 1 detected stress in 86.47% of the utterances and Model 2 detected stress in 72.43% of the utterances. Although initially trained on different phonetic dictionaries, in stage two both models detected stress on both the first syllable and second syllable. From the bar graph, we can see that Model 1 (trained on the first syllable in stage one) chooses the variant with second syllable stressed 44.76% across the utterances. Furthermore, while Model 2 (trained on the second syllable in stage one) predominantly chooses the variant with the second syllable stressed, it also chooses the variant with the first syllable stressed 20.96% across the utterances. This indicates that stress assignment is not restricted to the first syllable.

6 Conclusion

Thus, using an ASR system (Kaldi) I train finite state models to detect stress cues in a speech corpus and test the existing accounts of stress against results from the corpus. The results illustrate that the model detects stress cues on the initial syllable and the second syllable which strongly suggests that word-level stress assignment is not strictly word-initial. Thus, this agrees with Shaw's account (Proposal 2) over Hayes and Lahiri's account (Proposal 1).

Furthermore, the final results cannot be explained by the training probabilities alone. If the learning was entirely probabilistic, Model 1 and Model 2 would simply choose the variants they were trained on, i.e., the first syllable and second syllable respectively. However, since the models detected stress in positions they were not trained on, it illustrates that the models learned to detect stress cues in this study.

The final results do not definitively support the quantity sensitive account of stress, as proposed by Shaw. These models were trained without any cues about syllabification and that could be one reason why certain words were assigned primary stress on the first syllable in certain contexts and on the second syllable in another context. Word-medial clusters could either be a complex onset or it could be split up into the coda of a preceding syllable and the onset of the following syllable. Depending on the syllabification, this would lead to a syllable being classified as heavy or not and thus would affect the distribution of stress, as per the quantity-sensitive account.

As discussed earlier in Section 2.1, syllabification patterns vary across different strata in Bengali (Kar 2009) and it has been suggested that the distribution of stress could also vary across these strata (Chatterji 1921). One possibility could be that stress assignment is strictly word-initial in one stratum but quantity sensitive in another stratum.

However, the speech corpus Shruti has over 22,000 words and they are not tagged with the strata they belong in. Stratal information in Bengali is not readily available and future work in this area could include hand-tagging the lexicon with the stratal information and comparing the alignments of words within each stratum to test the distribution of stress for each stratum.

Thus, overall we can see that finite state models can be trained to detect prosodic and suprasegmental features in natural language and provide insight into phonological patterning. In the case of Bengali, the results suggest the stress assignment is not strictly word-initial, as previously suggested in the literature.

References

Allauzen, Cyril, Michael Riley, Johan Schalkwyk, Wojciech Skut, and Mehryar Mohri. 2007. OpenFst: a general and efficient weighted finite-state transducer library. In *International Conference* *on Implementation and Application of Automata*, ed. Jan Holub and Jan Žďárek, 11–23. Prague, Czech Republic.

Anderson, James Drummond. 1920. *A Manual of the Bengali Language*. London: Cambridge University Press.

Bykova, Evgeniia Mikhailovna. 1981. *The Bengali Language*. Moscow: Nauka Publishing House, Central Dept. of Oriental Literature. Original title: Bengal'skij Jazyk.

Chatterji, Suniti Kumar. 1921. Bengali phonetics. *Bulletin of the School of Oriental and African Studies* 2:1–25.

Das, Biswajit, Sandipan Mandal, and Pabitra Mitra. 2011. Bengali speech corpus for continuous automatic speech recognition system. In *Oriental COCOSDA 2011 : International Conference on Speech Database and Assessments*, 51–55. Taiwan: IEEE.

Das, Shyamal. 2001. Some Aspects of the Phonology of Tripura Bangla and Tripura Bangla English. Doctoral dissertation, CIEFL Hyderabad.

Ferguson, Charles A, and Munier Chowdhury. 1960. The phonemes of bengali. *Language* 22–59.

Fitzpatrick-Cole, Jennifer, and Aditi Lahiri. 1997. Focus, intonation and phrasing in Bengali and English. In *Intonation: Theory, Models and Applications*, ed. Antonis Botinis, 119–122. Athens, Greece: ISCA Archive.

Gildea, Daniel, and Daniel Jurafsky. 1995. Automatic induction of finite state transducers for simple phonological rules. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, 9–15. Association for Computational Linguistics.

Gordon, Matthew. 2011. Stress systems. In *The Handbook of Phonological Theory, Second Edition*, ed. Alan Yu John Goldsmith, Jason Riggle, 141–163. Wiley Online Library.

Gordon, Raymond G. 2005. *Ethnologue: Languages of the World*, volume 15. Dallas, TX: SIL international.

Goswami, Krishnapada. 1944. Linguistic notes on Chittagong Bengali. *Indian Linguistics: Journal of the Linguistic Society of India* 8:493–536.

Hasan, Md. Rashidul, Mustafa Jamil, Rabbani Golam, and Md. Saifur Rahman. 2004. Speaker identification using mel frequency cepstral coefficients. *Variations* 1:4.

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hayes, Bruce, and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language & Linguistic Theory* 9:47–96.

Hayes, Bruce Philip. 1980. A Metrical Theory of Stress Rules. Doctoral dissertation, Massachusetts Institute of Technology.

Johnson, Douglas C. 1970. Formal Aspects of Phonological Description. Doctoral dissertation, University of California, Berkley.

Kaplan, Ronald M, and Martin Kay. 1980. Regular models of phonological rule systems. *Computational linguistics* 20:331–378.

Kar, Somdev. 2009. The Syllable Structure of Bangla in Optimality Theory and its Application to the Analysis of Verbal Inflectional Paradigms in Distributed Morphology. Doctoral dissertation, Universität Tübingen.

Karttunen, Lauri. 1993. Finite-state constraints. In *The Last Phonological Rule: Relfections on Constraints and Derivations*, ed. John Goldsmith, 173–194. Chicago: University of Chicago.

Karttunen, Lauri. 2006. A finite-state approximation of Optimality Theory: the case of Finnish prosody. In *Advances in Natural Language Processing*, ed. Tapio Salakoski, Filip Ginter, Sampo Pyysalo, and Tapio Pahikkala, 4–15. Turku, Finland: Springer.

Kawasaki, Haruko, and Stefanie Shattuck-Hufnagel. 1988. Acoustic correlates of stress in four demarcative-stress languages. In *The Journal of the Acoustical Society of America*, volume 84, 98. Acoustical Society of America.

Khan, Sameer ud Dowla. 2011. Intonational Phonology and Focus Prosody of Bengali. Doctoral dissertation, University of California, Los Angeles.

Klaiman, Mimi H. 1987. Bengali. In *The World's Major Languages*, 490–513. Croon Helm.

Lahiri, Aditi, and Jennifer Fitzpatrick-Cole. 1999. Emphatic clitics and focus intonation in Bengali. In *Phrasal Phonology*, ed. René Kager and Wim Zonneveld, 119–144. Nijmegen: Nijmegen University Press.

Lewis, Paul M., Gary F. Simons, and Charles D. Fennig. 2015. *Ethnologue: Languages of the World*, volume 18. Dallas, TX: SIL International.

Liberman, Mark. 1975. The Intonational System of English. Doctoral dissertation, Massachusetts Institute of Technology.

Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8:249–336.

Mandal, Sandipan, Biswajit Das, Pabitra Mitra, and Anupam Basu. 2011. Developing Bengali speech corpus for phone recognizer using optimum text selection technique. In *International Conference on Asian Language Processing (IALP)*, 268–271. Penang, Malaysia: IEEE.

Michaels, Jennifer Marie, and Catherine E. Nelson. 2004. A preliminary investigation of intonation in East Bengali. Manuscript.

Mohri, Mehryar, Fernando Pereira, and Michael Riley. 2002. Weighted finite-state transducers in speech recognition. *Computer Speech & Language* 16:69–88.

Povey, Daniel, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, and Karel Vesely. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society. IEEE Catalog No.: CFP11SRW-USB.

Reynolds, Douglas A, and Richard C Rose. 1995. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing* 3:72–83.

Sapir, Edward, and Morris Swadesh. 1960. *Yana dictionary*. Berkeley, CA: University of California Press.

Selkirk, Elisabeth. 2007. Bengali intonation revisited: An optimality theoretic analysis in which focus stress prominence drives focus phrasing. In *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, ed. Chungmin Lee and Matthew Gordon, 215–244. Springer.

Shaw, Rameshwar. 1984. Stress-patterns in Bengali and Hindi: A comparative study. In *Papers in phonetics and phonology*, ed. Bando Bhimaji Rajpurohit. Mysore: Central Institute of Indian Languages.

Tryon, Darrell T. 1970. *An Introduction to Maranungku (Northern Australia)*. Canberra, Australia: Pacific Linguistics.

Weide, R. 1998. The CMU pronunciation dictionary, release 0.6.

Department of Linguistics
203 Morrill Hall
Cornell University
Ithaca, NY 14853
sb2295@cornell.edu