Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

A Quantitative Study of Focus Shift in Marathi

Preeti Rao[1], *Hansjörg Mixdorff*[2], *Ishan Deshpande*[1], *Niramay Sanghvi*[1], *Shruti Kshirsagar*[1]

[1] Department of Electrical Engineering, IIT Bombay, India
[2]Beuth University Berlin, Germany

prao@ee.iitb.ac.in, mixdorff@bht-berlin.de

Abstract

We study the effect of focus shift on prosodic features for Marathi, a major Indian language. In our analysis, we consider different focus locations and different focus widths. We report observations of fundamental frequency, intensity, and syllabic durations of constituent words of the utterance. *F0* is studied via the accent commands of the Fujisaki model. We contrast statements containing narrowly focused content words with broad focused statements. Contexts for narrowly focused items are either contrastive or non-contrastive. Our results show that narrow focus is marked by longer duration of the focused word, and a larger accent command in the focused word. Post-focal effects are observed for duration, intensity and *F0*. No differences were found between contrastive and non-contrastive focus.

Index Terms: prosody, focus, stress, Marathi

1. Introduction

Focus is an important functionality of prosody that is expressed by different degrees of prominence attributed to the words in an utterance. An understanding of its realization and perception is necessary for building a model of the prosody of any language. Studies across several languages have indicated that all languages draw from the following set of acoustic devices to signal prominence: duration, fundamental frequency (*F0*), intensity and spectral characteristics. In stress-accented languages, the lexically stressed syllable is the modified syllable in a focus word [1]. Two major aspects of focus are the width of focus (narrow, broad) and its location (position of the prominent word in the sentence). For American English [2], it was observed that the strongest indicators for discriminating the different focus conditions are duration (including subsequent pauses), mean and maximum *F0* (all measurements made across words) and intensity. Korean shows both narrow focus and post-focal effects that involve all the acoustic parameters [3]. In a recent study of German, perceptual prominence of a word was observed to be strongly correlated with *F0* transition, syllable duration, maximum intensity and mean harmonics-to-noise ratio [4]. Traditionally, studies of focus have considered the variation of pitch accents. Strong language dependence is observed as was shown, for example, in a comparison of Dutch with Italian [5].

Marathi, a language spoken predominantly in the Indian state of Maharashtra with its population of over 100 million, is a relatively poorly studied language as far as the prosody is concerned. However there exist a few studies on the prosody of Hindi [6, 7, 8, 9, 10]. Hindi and Marathi share numerous similarities with regard to the written word as well as pronunciation since they are both derived from Sanskrit, like several other Indo-Aryan languages. However Hindi is known to be influenced by Persian while Marathi has Dravidian influences in its phonetic inventory. Further, unlike in Hindi, phonological length opposition of vowels is not seen in present-day Marathi [11].

Several studies of Hindi prosody note that each non-final content word exhibits a *F0* rise with the L and H tones assigned to it. Rao and Srichand [12], in a more general study of *F0* variations in four Indian languages including Hindi and Marathi, observed that locally *F0* increases from the left to the right across every content word, while steadily falling across the sentence globally for declarative sentences. This observation was exploited by them for word segmentation of continuous speech in Hindi, as well as by others [13] who noted that *F0* plays a major role in demarcation of continuous speech in Hindi. This *F0* behavior - called a "hammock shape" by Pandey [14] who also observed it in Indian English - is important, since it may mean that the more significant *F0* changes would occur towards the end of the word. Hindi and Marathi are considered syllable-timed languages and lexical stress per se is not a well understood aspect. However there have been some commentaries that assign lexical stress based on syllable weight in Hindi [15] and Marathi [16] although the acoustic correlates of stress are unknown. It is of interest to explore the interaction of focus with the hypothesized syllabic stress on the one hand and the characteristic hammock *F0* shape on all non-final content words on the other hand. A related study [6] on contrastive focus in Hindi found an increase in duration of the stressed syllable as well as increased *F0* rise across the word that came as much from lowering the L tone as raising the H tone in the hammock shape. Puri [7] found the main acoustic correlates of focus in Hindi by bilingual speakers (of Indian English) to include increased duration as well as an *F0* excursion on the focused element and post-focal reduction in duration, amplitude and *F0* excursion. No amplitude increase was observed on the focused word. Post-focal compression of pitch range was also noted previously [8, 9, 10].

In this study, we aim to find the prosodic cues of focus in Marathi. This is, to the best of our knowledge, amongst the first works on the topic. Therefore we will follow the methods that have been employed in studying other languages. We consider all the three dimensions of *F0*, duration and intensity that have been traditionally studied across several of the world's languages in the context of focus. We investigate the features across the sentence, so as to capture local as well as global effects. *F0* behaviour is further studied via its parameterization by the Fujisaki model. The eventual goal is to establish the features that are the most significant indicators of focus, in terms of acoustic features as well as their perception given the peculiarities of the language including the mandatory *F0* variation across a word and hypothesized lexical stress rules. We next describe the experimental setup. Then we introduce the Fujisaki model. This is followed by the observations, which are reported in three sub-parts, one each for duration, intensity and *F0*.

 doi: 10.21437/SpeechProsody.2016-192

2. Speech Material and Method of Analysis

2.1 Data Set

Target	Prompt
Amol aai barobar bolat hota *(Amol/mother/with /talking/was)* Amol was talking with mother	*Tumhi kay mahnalat?* What did you say? *Kon aai barobar bolat hota?* Who was talking with mother? *Amol konabarobar bolat hota?* With whom was Amol talking? *Amol aai barobar kay karat hota?* What was Amol doing with mother?
Nahi…Amol aai barobar bolat hota No… Amol was talking with mother	*Rohit aai barobar bolat hota ka?* Was Rohit talking with mother? *Amol bhavabarobar bolat hota ka?* Was Amol talking with brother? *Amol aai barobar khelat hota ka?* Was Amol playing with mother?

Table 1: *Target and prompt texts in Romanized script with the English translation.*

Marathi is an SOV language with flexible word order. Focus can be signaled by change in syntactic word order and/or by a morpheme. Given this, it was important to confirm that eliciting varying prominence via purely prosodic features - by constraining the target sentence - came easily and naturally to native Marathi speakers. The selected target sentence, as seen in Table 1, has three critical words: subject (*Amol*), object (*aai*) and verb (*bolat*), and thus allows the study of both focus type and location. Dhongde [16] mentions that accent is not distinctive in Marathi, but provides a tentative set of stress rules based on the syllable weight in multi-syllabic words. Using the rules given by Dhongde, we hypothesize that '*mol*' in *Amol*, and '*bo*' in *bolat*, and are the lexically "stressed" syllables, and hence potential candidates for receiving emphasis due to focus (henceforth referred to as stressed syllables). This adds another dimension to our study, since the position of the stressed syllable within the word is different in two of the critical words.

Data from a total of 20 native speakers of standard Marathi were collected. All were young adults studying or working in Mumbai. Eventually, we used the data of 12 speakers (6 male and 6 female) for our analyses based on a verification procedure described later. Seven pre-recorded questions by a different native Marathi speaker were used as prompts to elicit appropriately focused responses as shown in Table 1 for narrowly focused non-contrastive and contrastive statements, and one for the broadly focused. Two instances were elicited of each of the target forms providing 14 utterances per speaker. The target utterance for the contrastive form includes the prefix *Nahi* which the speakers were instructed to articulate silently. As an introduction to the task, examples of questions and responses by a native Marathi speaker corresponding to a different SOV sentence were played out to the test speakers without further explanation. We recorded our speakers with a high quality microphone in a quiet room at 16 kHz sampling rate. The data collection was followed by perceptual verification. Two listeners (native Marathi speakers who did not participate in the recording) were presented each recorded statement over headphones and asked to identify the location and width of focus to verify that it matched the intended location/width. We retained only those speakers for analyses who had all 14 utterances pass the verification with both listeners. This led to a dataset of 168 utterances (12 speakers x 7 focus conditions x 2 instances).

We observed that while most focus conditions were reliably discriminated across the 12 speakers, this was not true of the contrast-distinction which remained around chance. The listeners also agreed that the recordings sounded natural.

2.2 Acoustic Measurements

All utterances were manually aligned with the help of the spectrogram and waveform views in PRAAT [17] keeping in mind the phones constituting the syllable (V, CV or CVC corresponding to our critical words: *A-mol, aai, bo-lat*). We explicitly segmented speech pauses of duration over 100 ms. Silences of shorter durations were merged with the preceding syllable. Intensity and *F0* contours sampled at 10 ms intervals were extracted with PRAAT. Measurements include mean and maximum intensity of every syllable, the mean and maximum intensity of each word, and the duration of each syllable. We also measured *F0* in semitones with respect to utterance mean *F0* for the word-level maximum, minimum and *F0* span. For each word, we find the word-level minimum and the time at which it occurs. To ensure that the measured word-level maximum is part of an *F0* rise, we scan for it in the time span following the observed minimum. The word-level minimum and maximum constitute *F0* min and *F0* max respectively with the difference providing the *F0* span. Apart from the perceptual similarity observed earlier, preliminary inspection of individual speakers' data did not reveal any differences in the acoustics between the non-contrastive and contrastive forms. We therefore pooled the data for each of the three focus location forms to get four utterances each per speaker. Only the broad focus was represented by two utterances per speaker.

2.3 Extraction of Fujisaki Model Parameters

Apart from measurements on the raw *F0* contours, we approximate each contour with the Fujisaki model by superimposing three components: A constant base frequency *Fb*, exponentially decaying phrase components which are the responses to the phrase commands and accent components which are the smoothed responses to the accent commands. In Fig. 3, panel (1), we see an example of *F0* contour decomposition for the non-contrastive version with focus on *Amol* uttered by male speaker AK. The natural *F0* contour is indicated by +++ signs. It is modelled using one phrase component triggered by an impulse-wise phrase command with magnitude *Ap* of 0.36 at t=-0.11 s and one box-shaped accent command with *Aa*=0.51 and onset time *T1* at t=0.54 s and offset time *T2* at t=0.75 s. The modelled contour is indicated by the solid line which approximates the natural contour closely. *Fb* of 109 Hz is indicated by the horizontal dotted line. We extract the Fujisaki model parameters underlying the *F0* contour by applying the method presented in [18], and compare the accent commands.

3. Results of Analysis

We use the speaker's broad focus statement as the baseline and compare the narrow focus statements against it.

3.1 Duration

We converted the log duration of each speaker-syllable to its utterance-level normalized Z-score to compensate for speech

rate. Next we calculate log duration distributions for each syllable and word across speakers and utterances for the different focus conditions (broad, narrow, pre-, post-). Fig. 1 shows the obtained box-plots for duration and Table 2 the 1-way ANOVA analyses for all measured parameters.

We observe in Fig. 1 that focus type is clearly cued by word duration for all focus locations. Word duration is extended in focus while it is reduced post-focally, all with respect to the neutral focus. Pre-focus durations (of *Amol* and *aai*) are not discriminated from neutral focus. Fig. 1 also captures the interesting observation that it is the stressed syllable (i.e. final in *Amol*, and penultimate in *bolat*) that undergoes elongation in focus and reduction when in post-focus. An additional observation was that of pre-focal pauses. We detected 47 inter-word pauses across the 12 speakers' data. Of these 37 were pre-focal (26 before *aai*, and 11 preceding *bolat*). The remaining ones appeared after the word *Amol* in narrow or broad focus.

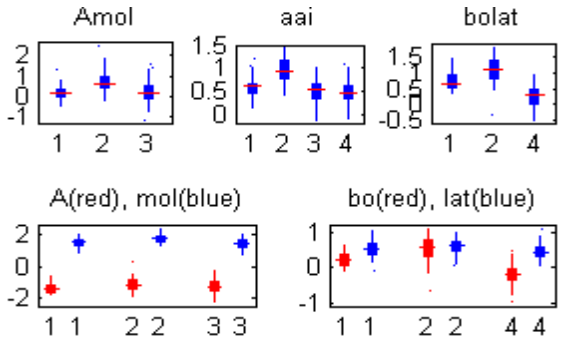

Figure 1: *Normalized log duration distributions: words (top), syllables (bottom). (1:Broad focus, 2:on narrow focus, 3:pre-focally, 4:post-focally)*

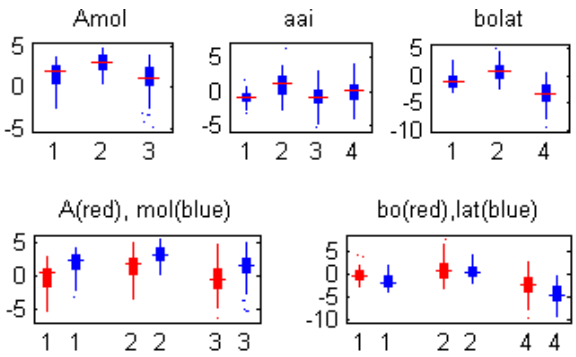

Figure 2: *Mean intensity distributions: words (top), syllables (bottom). (1:Broad focus, 2:on narrow focus, 3:pre-focally, 4:post-focally)*

3.2 Intensity

To eliminate inter-speaker variability, we normalize all measurements of intensity to the mean intensity of the corresponding utterance. We observe in Fig. 2 that focus shift does affect the mean intensity of the word relative to broad focus. However the only consistent trend observed across speakers was the post-focal decrease in intensity on the verb (*bolat*), as also borne out by mean intensity F values in Table 2. In the disyllabic words, it was observed that both syllables are similarly affected as seen in Fig. 2.

Paramt	**On focus**			**Post focus**	
	Amol	*aai*	*bolat*	*aai*	*bolat*
Logdur	20.3 **(0)**	28.2**(0)**	17.4**(0)**	3.2**(0.07)**	43.6**(0)**
I (max)	11**(0.001)**	17.3**(0)**	9.8**(0.002)**	0.6 **(0.43)**	25.2**(0)**
I (mean)	22.1**(0)**	18.3**(0)**	16.6**(0)**	3.8**(0.05)**	33.2**(0)**
F0 (Max)	10**(0.002)**	62.1**(0)**	17.5**(0)**	0.4**(0.5)**	206.0**(0)**
F0(Span)	0.43**(0.5)**	50.5**(0)**	16.1**(0)**	21.0**(0)**	201.6**(0)**

Table 2: *F values (p values in parentheses) from 1-way ANOVA (p=0.05) with reference to broad focus; p values<0.001 set to 0.*

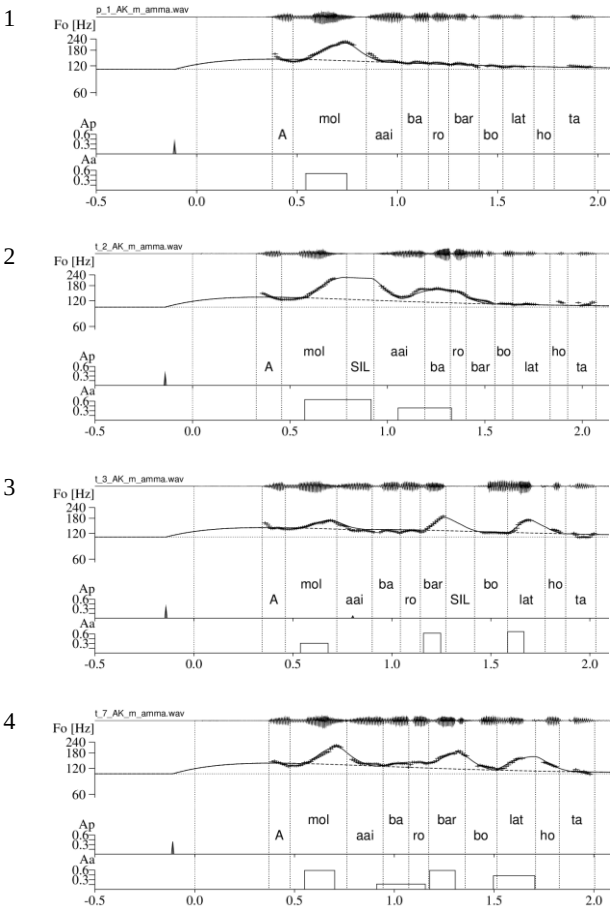

Figure 3: *Speech waveform, F0 contour and underlying Fujisaki model parameters for 'Amol aai barobar bolat hota' - (top to bottom) foci: (1) narrow on Amol, (2) narrow on aai, (3) narrow on bolat, and (4) broad.*

3.3 *F0*

Fig. 3 presents results of the Fujisaki analysis for speaker AK's target sentence for the following conditions: (1) Non-contrastive focus on *Amol*, (2) *aai*, (3) *bolat*, and (4) broad focus. Each of the four panels displays, from the top to the bottom: the speech waveform, the *F0* contour, and the underlying phrase and accent commands. The syllable segmentation is indicated by the dotted vertical lines. Marathi syllable texts are provided in a Romanized transcription.

As can be seen, *F0* contours differ clearly for all four conditions. Except for post-focal words we indeed observe a well-defined hammock-like shape as reported by [14]. This

shape results from accent commands which are aligned with the right edge of the constituent words. In the case of *aai* narrowly focused, the accent command actually extends into *barobar*. This continuation is even more striking in the broad focus case: A relatively weak accent command on *aai* is further boosted by a second command associated with the last syllable of *barobar*. This indicates that *aai barobar* is treated as a prosodic unit. AK, like many other subjects, further emphasizes narrow focus on *aai* and *bolat*, respectively, by introducing a short pause before the focused constituents. Further, as may be expected from the language characteristic hammock shaping at word level, the significant pitch increases in the disyllabic words were observed to take effect in the final syllables (*-lat, -mol*) as measured by syllable mean pitch.

If we compare the three narrowly focused conditions, the main difference is actually seen on the accent commands associated with the post-focal items. With *bolat* in focus the associated accent command is boosted, whereas the command becomes deleted when narrow focus is placed on *Amol*. In contrast, the accent command aligned with *Amol* is never suppressed. This suggests that narrow focus only has a marginal effect on pre-focal items, but suppresses *F0* gestures on the post-focal ones.

Table 3 displays means of *Aa* for all syllables in the four focus cases. The non-existence of an accent command associated is taken into account with an *Aa* of zero. As can be seen, *Aa* is low for post-focal items (set in grey). Independent samples Kruskal-Wallis test shows that *Aa* assigned with the critical words is significantly different for *aai, bolat* depending on the focus condition ($p < 0.001$) whereas the significance is lower for *Amol* ($p < 0.007$).

From the 1-way ANOVA results on *F0* measurements on the raw contours presented in Table 2, we note that the maximum *F0* and the *F0* span are significantly higher for narrow focus at object and verb locations with reference to neutral focus. The *F0* cues are not so clear for the S focus (*Amol*). Post-focus on the V (*bolat*) is cued by both *F0* measures, with significant decrease in value compared with neutral focus. Post-focus on the O (*aai*), on the other hand, is cued only with a decrease in pitch span. These observations support those on the accent command presented earlier.

syll.	*Amol*	*aai*	*bolat*	Broad
A	0.00	0.00	0.00	0.00
mol	**0.46**	0.45	0.36	0.45
aai	0.12	**0.37**	0.12	0.18
ba	0.07	0.33	0.16	0.18
ro	0.07	0.26	0.26	0.28
bar	0.08	0.18	0.48	0.50
bo	0.05	0.19	0.22	0.29
lat	0.03	0.02	**0.49**	0.38
ho	0.02	0.03	0.04	0.14
ta	0.00	0.00	0.00	0.00

Table 3: *Accent command amplitudes Aa for the syllables of the target sentence, value for focused word set in bold type.*

4. Discussion and Conclusions

Observations from prominence-eliciting experiments have been presented as part of a larger study on prosody and information structure in Marathi. Duration, intensity and *F0* have been measured at word and syllable level for the target sentence in SOV form. The *F0* measurements also involve accent parameters obtained from Fujisaki model fitting of the natural contours. Statistical analyses indicate that all the word-level acoustic parameters (duration, intensity mean, *F0* max and span) cue on-focus in object and verb. For the subject, the same holds except that *F0* cues are not so discriminative for on-focus with respect to broad focus probably due to topic marker effect. In agreement with this, Fujisaki model accent command amplitude, expected to be more robust than *F0* max, shows a clear increase with narrow focus in the object and verb locations but not in subject focus. Pre-focus is not distinguished from broad focus in any of the attributes. Post-focal compression is most clearly observed in duration, intensity and *F0* measurements in the verb location.

In the two disyllabic words (*Amol, bolat*), it is the stressed syllables that are affected for duration. However, it is the word-final syllables that are affected for *F0* increase as seen from the obtained alignment of the accent command. Thus, the regular hammock shape characteristic of content words is emphasized further with *F0* increase on the final syllable when the word is in focus.

An interesting observation is the case of *aai* which forms a unit with the following function word *barobar* so much so that the strongest *F0* excursion appears on the final syllable *bar* in broad and pre-focus conditions. Here the *aai* segment shows a variety of *F0* realizations, most commonly a slight lowering of *F0*. When *aai* is in narrow focus however, there is a well-defined hammock on the focus word with *barobar* showing a non-increasing *F0* including a fall over one or more syllables. This behaviour is also captured by the changing alignment of the accent command with respect to the syllable *aai*. Focus on the preposition *barobar* was not investigated here and is an interesting question.

Our observations are not very different from what has been seen for other languages, including Hindi [6, 7], where focus has been seen to be marked by increased *F0* span and elongation of the focused word. We find that post-focal reduction of duration and *F0* span are very conspicuous too (again as observed in Hindi [8, 9]), especially on the verb (*bolat*). Again, it is interesting that it is not our hypothesized stressed syllable in *bolat* (*bo*) where the significant changes of *F0* occur. The change occurs in the second syllable even when the first syllable is the stressed syllable according to the syllable weight based rule. We thus observe the trend of the stressed syllable to be lengthened or shortened when in or out of narrow focus respectively whereas intensity, if utilized by the speaker, is affected across all syllables of the word.

Finally, in our acoustic analysis we identified features that are modified under focus shift. In order to examine their relative perceptual contributions, we intend to carry out perception experiments using resynthesized speech. Thanks to the Fujisaki modeling, *F0* contour modifications are straightforwardly realized. Applying PSOLA based resynthesis will provide direct control of durations and intensity. Hence we will be able to examine more quantitative relationships between features and perceived changes in focus.

5. Acknowledgements

This work was supported by DFG international collaboration grant Mi 625 funding mutual visits by Mixdorff and Rao.

6. References

[1] Campbell, N. and Beckman, M., "Stress, prominence, and spectral tilt", In: *Intonation: Theory, models and applications*, 1997.

[2] Breen, M., Fedorenko, E., Wagner, M. and Gibson, E., "Acoustic Correlates of Information Structure", Language and Cognitive Processes 25.7, pp. 1044-1098, 2010.

[3] Lee, Y. C. and Xu, Y., "Phonetic realization of contrastive focus in Korean", In: Proc. Speech Prosody 2010, Illinois, 2010.

[4] Mixdorff, H., Cossio-Mercado, C., Hönemann, A., Gurlekian, J., Evin, D. and Torres, H., "Acoustic Correlates of Perceived Syllable Prominence in German", In: Proc. Sixteenth Annual Conference of the International Speech Communication Association, Dresden, 2015.

[5] Swerts, M., Krahmer, E. and Avesani, C., "Prosodic marking of information status in Dutch and Italian: A comparative analysis", Journal of Phonetics, vol. 30, no. 4, pp. 629-654, 2002.

[6] Genzel, S. and Kügler, F., "The prosodic expression of contrast in Hindi", In: Proc. 5th International Conference of Speech Prosody, Chicago, USA, pp. 1-4, May 2010.

[7] Puri, V., "Intonation in Indian English and Hindi late and simultaneous bilinguals", Ph.D. thesis, University of Illinois at Urbana-Champaign, 2013.

[8] Féry, C., "Indian Languages as Intonational 'Phrase Languages", In: *Problematizing language studies*, 2010, pp. 288–312.

[9] Harnsberger, J. D., "Towards an intonational phonology of Hindi", In: Proc. Fifth Conference on Laboratory Phonology, Northwestern University, 1996.

[10] Patil, U., Kentner, G., Gollrad, A., Kügler, F., Féry, C., and Vasishth, S., "Focus, word order and intonation in Hindi," Journal of South Asian Linguistics, vol. 1, no. 1, pp. 53–67, 2008.

[11] Yardi, V., "Teaching English pure vowels to Marathi learners: some suggestions", ELT Journal. 4, pp. 303-307, 1978.

[12] Rao, G. and Srichand, J., "Word Boundary Detection using Pitch Variations", In: Proc. International Conference on Spoken Language Processing, Philadelphia, 1996.

[13] Rajendran, S. and Yegnanarayana, B., "Word boundary hypothesization for continuous speech in Hindi based on F0 patterns", Speech Communication, 18(1), pp. 21-46, 1996.

[14] Pandey, P., "Indian English Prosody", In: Leitner, G., Hashim, A., and Wolf, H. (Eds.), *Communicating with Asia*, Cambridge University Press, Cambridge, 2015.

[15] Dyrud, L. O., "Hindi-Urdu: stress accent or non-stress accent?", Ph.D. thesis, University of North Dakota, 2001.

[16] Dhongde, R. V. and Wali, K., "Marathi", Vol. 13, In: *John Benjamins Publishing*, 2009.

[17] Boersma, P., "PRAAT, a system for doing phonetics by computer", Glot International 5, pp. 341-345, 2001.

[18] Mixdorff H., "A novel approach to the fully automatic extraction of Fujisaki model parameters", In: Proc. ICASSP 2000, vol. 3, pp. 1281-1284, Istanbul, Turkey, 2000.