# Significance of Duration in the Prosodic Analysis of Assamese

*D. Govind*[1]*, S. Mahanta*[2] *and S. R. Mahadeva Prasanna*[1]

[1]Department of Electronics & Electrical Engineering
[2]Department of Humanities and Social Sciences
Indian Institute of Technology Guwahati, Assam, India
govinddmenon@gmail.com, {smahanta, prasanna} @ iitg.ernet.in

## Abstract

The objective of the present work is to demonstrate the significance of duration in the context of phonological *Focus* of Assamese. Focus refers to that part of sentence which expresses assertion, putting more emphasis on that part of the sentence which introduces new information. The present work considers subject object verb (SOV) type declarative sentences in wide, object and subject focus cases for the study. Speech data was collected from native Assamese speakers in all the three types of focus. Manual duration analysis was carried for all the speech data. It was observed that compared to wide focus, the duration reduces in the object and subject focus cases. Even though the overall duration reduction in object and subject focuses is nearly same, the amount of reduction is different for subject (S), object (O) and verb (V) parts. The duration modification of wide focus speech according to the duration modification factors of either object or subject focus confirms that duration indeed influences the realization of focus.

**Index Terms**: Focus, duration modification, SOV, wide focus, object focus and subject focus

## 1. Introduction

Focus refers to that part of the sentence which relays more information on the important part of a sentence. Focus has relation to the meaning conveyed in a sentence (semantics). Change in the focus association of focus sensitive particles like *only* leads to distinctly different interpretations of the same sentence (sentence with the same word order). Focus information tries to give prominence to *new information* (new words) while old words (or given information) are not accented. The accented word(s) forms the focus domain. However, not all of the words in a focus domain need be accented. The English expression, *In the BIG house* could be a reply to *Is that where you live?*, The focused element here could be *BIG* and the rest of the phrase indicating the house or the building could be interpreted in the context of the discussion. By contrast, *In the BIG HOUSE*, which might be spoken in response to *Where would you have a party in that block of houses?*, that is in a situation where the respondent would be able to exercise an option from a group of houses, the focused constituent would be *BIG HOUSE*. Ladd [1] referred to the difference between the two responses and their focus constituents as the difference between *broad focus*, i.e. where the whole expression bears focus, and *narrow focus*, any focus constituent which is smaller than the whole expression.

Most importantly in the context of this paper, focus also has a significant role to play in the prosodic aspects of phonology. Focus has important implications for suprasegmental phonology (intonation, stress and duration), as specific intonational tunes are encoded in the grammar to encode focus. In most languages, speakers can use pitch accents on particular syllables to provide focus information in a particular utterance. Studies on focus and its ramifications on syntax and prosody have pointed out different types of focus phenomena. As in the example in the previous paragraph, the widely attested type is the question-answer type of focus and is called *presentational focus* [2–4] or *information focus* [5]. In this paper, an experiment conducted on Assamese will present certain aspects of prosody in presentational focus which may be important for speech synthesis. In particular, the duration aspect of prosody is considered.

In human-human speech communication, focus can be realized and comprehended in an effortless manner. However, it is a scientific curiosity and also signal processing challenge to mimic the same on a digital machine. For this to happen, first we need to understand how it is manifested in the speech signal. This is because, from signal point of view, focus is such a subtle information, there may not be very significant changes from the focus to non-focus part of the speech signals. Therefore a careful analysis and interpretation is required for the information present at various levels. The present work focuses on the duration aspect to see how focus affects the duration information. From the analysis point of view, we would like to know whether there is any modification in the duration from the wide focus to object and subject focus cases. If so, how much it is and also how it is distributed? Can we modify the wide focus speech to incorporate the change in duration information according to the target focus, namely, object and subject focus? Will the duration modified signal sound like object or subject focus speech?

The present work conducted an experiment on native speakers of Assamese speaking an Eastern variety of the language. For the purpose of analysis we took 2 declarative sentences controlled for the number of syllables. Each sentence was produced in three different types: SOV (wide focus), SOV (subject focus) and SOV (object focus). All these sentences were produced in response to a scripted question and answer pattern written in the Assamese orthography. Four speakers were recorded for this experiment. The speech signals were processed in praat to identify the duration of each of the subject (S), object (O) and verb (V) regions [6]. The durations were averaged across the speakers for each of the wide, object and subject focus cases. The information from this averaging was interpreted to find out the effect of focus on duration part. Later the same information is used for duration modification to justify the need for such a study in speech synthesis framework using zero frequency filtering based prosody modification.

The organization of the paper is as follows: Section 2 explains the effect of duration in subject and object focus in Assamese. Section 3 describes the epoch based duration modi-

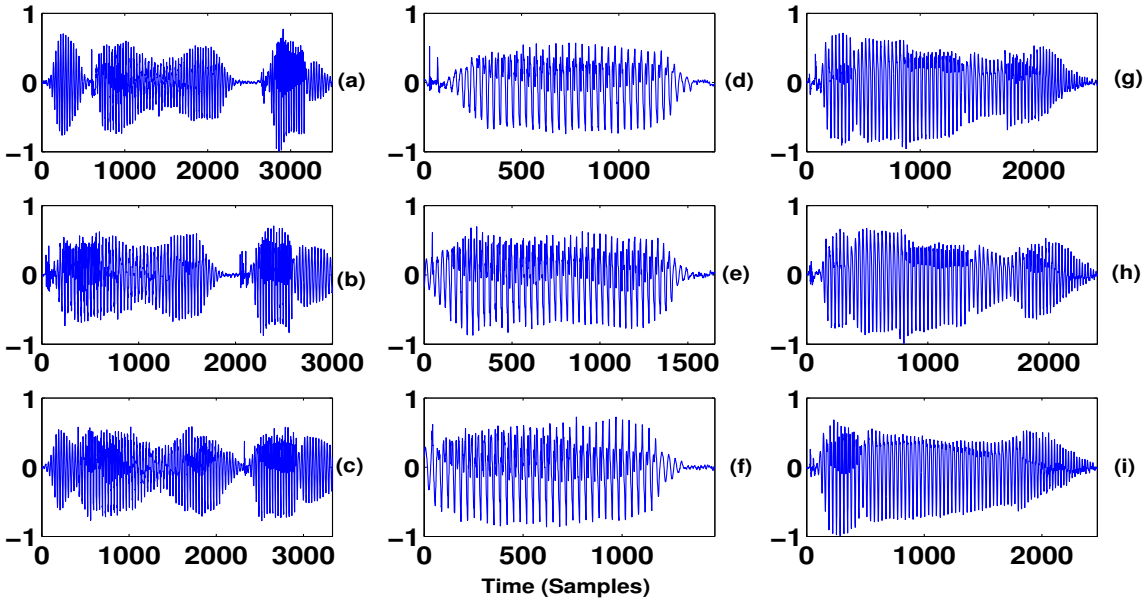


Figure 1: Duration of Subject, Object and Verb parts in wide focus, object focus and subject focus. ((a)-(c)) subject part, ((d)-(f)) object part and ((g)-(i)) verb part of wide, object and subject focussed Assamese sentence, respectively.

fication and Section 4 explains the procedure to convert SOV subject or SOV object focus from SOV wide focus by duration modification. The subjective study performed to evaluate the perceptual effectiveness of synthesized focus information is given in Section 5 and finally Section 6 concludes with scope for future work.

## 2. Effect of duration in subject and object focus in Assamese

The effect of duration in subject and object focus is analyzed with respect to the wide focus case. Figure 1 shows the subject, object and verb parts of a wide, object and verb focussed assamese data with the SOV word order. As we can observe that the duration of the object, subject and verb parts of the wide, subject and object focus are significantly different indicating the role of duration in the phonological focus of the Assamese language. The Figure 1 indicates that while applying focus to a subject or object part of a sentence in SOV, the duration of the other parts are affected. For instance, the subject part of the object focus shown in Figure 1(b) is found to be compressed and the duration of the object part (Figure 1(e)) is found to be equal or slightly increased. To statistically analyze the effect of focus in assamese, a data set of 4 speakers and 2 sentences is selected from 10 speaker, 10 sentence database. The following subsection 2.1 gives the details of the database used and subsection 2 provides the duration analysis performed across the data set selected for the present work.

### 2.1. Database

The raw database used for this research consists of data collected for a linguistic investigation of the phonological and phonetic markers of focus in Assamese. In order to investigate the relationship between word order and prosody, the data was collected from 10 native speakers of Assamese speaking an Eastern variety of the language. The speakers were all female and they were between 20-22 years of age. There are 10 declarative sentences controlled for the number of syllables. Each sentence was produced in six different types: SOV (wide focus), SOV (subject focus), SOV (object focus) OSV (wide focus), OSV (object focus), OSV (subject focus). All these sentences were produced in response to a scripted question and answer pattern written in the Assamese orthography. The recorder used for the experiment was a PMD Marantz 670 and a Sennheiser e914 microphone. The recorder's settings were fixed at 16 bit with a sampling rate of 44.1KHz.

### 2.2. Duration Analysis

The duration in terms of samples is estimated using the manual marking of the subject, object and verb parts of the utterances in Praat. The average duration for subject, object and verb parts of all the 4 speakers of two sentences are computed. From Table 1, it has to be observed that the overall duration of the speech with object and subject focus are different as compared to the wide focus case. From Table 1 it also has to be observed that the duration characteristics are different for subject, object and verb parts of the subject and object focused utterances. For instance, the subject part of the object focus utterance is time compressed as compared to the wide focus and duration of object and verb parts remain more or less same as that of the wide focus case. Even though there is no significant duration variation in the verb part of the subject focus for the sentence 1 but there is significant time compression is observed in the verb part in subject focus for sentence 2.

Table 2 provides the scaling factors derived from the average obtained for sentence 1 and sentence 2 in Table 1. This scaling factor is used to convert from a sentence in one focus to another by duration modification. The scaling factors shown in Table 2 are obtained by dividing average duration of the SOV part having the desired focus with corresponding SOV

Table 1: Average Duration (samples) of subject, object and verb parts of wide, object and subject focus utterances.

| Focus | Subject | Object | Verb | Average |
|---|---|---|---|---|
| Sentence 1 | | | | |
| Wide focus | 3779 | 2626 | 2001 | 8407 |
| Object Focus | 3164 | 2634 | 1991 | 7790 |
| Subject focus | 3150 | 2386 | 2057 | 7594 |
| Sentence 2 | | | | |
| Wide focus | 3369 | 2092 | 2809 | 8270 |
| Object Focus | 2967 | 2012 | 2766 | 7746 |
| Subject focus | 3194 | 2028 | 2556 | 7780 |

Table 2: The scaling factors derived from the average duration of subject, Object and verb parts of wide, object and subject focus utterances.

| Focus | Subject | Object | Verb | Average |
|---|---|---|---|---|
| Object Focus | 0.86 | 0.99 | 0.99 | 0.93 |
| Subject focus | 0.89 | 0.94 | 0.96 | 0.92 |

parts with the wide focus. These duration scaling factors are set for SOV parts in the wide focus utterance to synthesize utterance in the desired focus. To synthesize a good quality duration modified speech with reduced perceptual distortion, an epoch based duration modification algorithm is used. The following section explains the steps involved in the epoch based duration modification

## 3. Epoch based duration modification

The algorithmic steps involved in the epoch based prosody modification in [7] are used to propose more accurate zero frequency filtering based fast prosody modification in [8]. The steps in duration modification are as follows:

1. ***Finding the accurate pitch marks***
   The accurate epoch locations estimated using the zero frequency filtering method is used as the pitch marks for duration modfication [8, 9].
2. ***Deriving the synthesis pitch marks***
   The synthesis pitch marks have to be derived according to desired duration modification factors. This can be obtained from the analysis pitch marks. An epoch interval plot is derived by finding the intervals between successive epoch locations (Analysis pitch marks). This epoch interval corresponds to the instantaneous pitch period. To modify duration, the epoch interval plot obtained is interpolated and resampled according to the duration modification factor. Starting from a point the new locations are derived from the modified epoch interval plot (interpolated and resampled original epoch interval). These new locations are used as the synthesis pitch marks for the waveform reconstruction.
3. ***Waveform reconstruction***
   To reconstruct the duration modified speech, the original epoch locations that corresponds to the modified epoch locations are found first. The waveform samples in the original epoch intervals are copied to the corresponding modified epoch locations. In this way some of the epoch intervals are repeated (in case of increase in duration) and some of the epoch intervals are deleted (decrease in duration) in the duration modified speech.

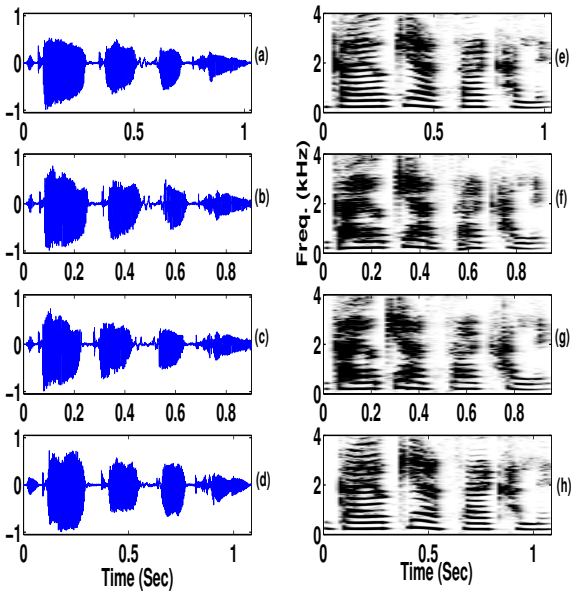


Figure 2: Synthesizing the object focus from wide focus by duration modification. (a) Speech waveform of an utterance with wide focus, (b) synthesized object focus by gross level duration modification, (c) synthesized object focus by the duration modification of subject, object and verb parts independently and (d) target speech waveform for the object focus.

## 4. Synthesizing Subject and Object focus from Wide Focus

To understand the significance of duration characteristics in perceiving subject and object focused SOV utterances, the wide focused SOV utterance is subjected to duration modification. The subject and object focus in a given wide focus utterance is synthesized by time scaling the overall duration of the wide focus speech according to the respective duration modification factor. Unlike Table 1 duration characteristics are different for subject, object and verb parts of subject and object focus utterances as compared to wide focus speech. Therefore different modification factors have to be set for subject, object and verb parts of the wide focus speech to synthesize the utterance in subject and object focus. Figure 2 plots the synthesized speech with object focus by gross level duration modification and duration modification of subject, object and verb parts independently. The spectrogram representations given in Figure 2 indicate that there are no perceptual distortion present in both the synthesized cases. Figure 3 plots the synthesized subject focus speech from the wide focus utterance. The spectrograms shown in Figures 2 and 3 indicate that there are no spectral distortions present in the synthesized speech compared with the source speech in wide focus.

To evaluate the effectiveness of the duration modification in synthesizing the speech in subject and object focus, a comparative subjective test is carried out. The following section describes the subjective study conducted.

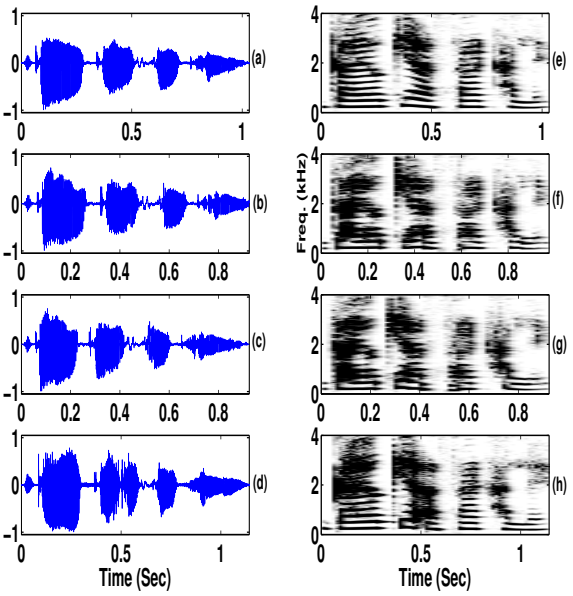


Figure 3: Synthesizing subject focus from wide focus by duration modification. (a) Speech waveform of an utterance with wide focus, (b) synthesized subject focus by gross level duration modification, (c) synthesized subject focus by the duration modification of subject, object and verb parts independently and (d) target speech waveform for the utterance in subject focus.

Table 3: Ranking used in perceptual test for CMOS.

| Rating | Description for evaluating synthesized speech |
|---|---|
| 1 | sounds exactly like source |
| 2 | sounds slightly different from source |
| 3 | sounds like from target focus |
| 4 | sounds sounds more like the target focus |
| 5 | sounds exactly like target |

## 5. Subjective Evaluations

All the files used for the subjective study are down sampled to 8 kHz from the 44.1 kHz which is used to originally record the speech. 10 research scholars who speak Assamese as a first language were participated in the subjective evaluation. The subjects were asked to provide a comparative mean opinion score (CMOS) for the synthesized files by comparing the source speech file in the wide focus and the speech file in the target focus. The filenames of duration modified files by gross level and SOV level modification are coded and randomized before presenting to the subjects. The subjects have to listen the wide focus speech (source) and speech in target focus and asked to provide CMOS in five point scale for the synthesized files. The significance of each score is described in Table 3. A total of 32 files used for the evaluation include 16 ($2\times2\times4$) synthesized files and 16 original files.

Table 4 shows the CMOS obtained for the synthesized SOV (object focus) and SOV (subject focus) by comparing the corresponding wide focus, target original SOV (object focus) and SOV (subject focus). The CMOS shows the effect of duration in the SOV (object focus) and SOV (subject focus). CMOS also indicate that time scaling the subject, object and verb parts separately provide more effective subject and object focus synthesis by duration modification.

Table 4: CMOS for the synthesized SOVO and SOVS.

| Focus | Gross Dur mod. | SOV Dur Mod. |
|---|---|---|
| Object focus | 1.8 | 2.85 |
| Subject focus | 3.1 | 3 |

## 6. Summary and conclusion

The present work demonstrated the effect of duration in the prosody of focus in Assamese sentences with SOV word order. The duration analysis shows that even though the sentence duration differences of the SOV (object focus) and SOV (subject focus) are nearly same with respect to the duration of SOV (wide focus) are nearly same, the durations of subject, object and verb parts are different for SOV (object focus) and SOV (subject focus). CMOS obtained demonstrates the significance of this duration information in converting a wide focus sentence into subject and object focus sentences.

The $F_0$ parameters have to be studied and incorporated along with the durational information to improve the effectiveness SOV (object and subject focus) conversion from SOV (wide focus).

## 7. Acknowledgements

This work is a part of ongoing UKIERI project (2007-2011) titled, Study of *Source Features for Speech Synthesis and Speaker Recognition* between IIT Guwahati, IIIT Hyderabad and CSTR, University of Edinburgh, UK.

## 8. References

[1] D. Ladd, *The Structure of Intonational Meaning: Evidence from English*, Bloomington and London:, Eds. Indiana University Press, 1980.

[2] E. Selkirk, "Contrastive focus vs. presentational focus: Prosodic evidence from the right node raising in english," in *Proc. Speech Prosody 2002*, 2002.

[3] M. L. Zubizarreta, *Prosody, Focus, and Word Order. Cambridge*, M. Cambridge, Ed. MIT Press., 1998.

[4] E. Vallduvi, *The Informational Component*. New York: Garland., 1992.

[5] K. Katalin, "Identificational focus versus information focus," *Linguistic Society of America, Language*, vol. 74, no. 2, pp. 245–273, 1998.

[6] P. Boersma and D. Weenink, "Praat 5.2.22: A software for doing phonetics," in *http://www.fon.hum.uva.nl/praat/*, 2011.

[7] K. S. Rao and B. Yegnanarayana, "Prosody modification using instants of significant excitation," *IEEE Trans. Audio, Speech and Language Processing*, vol. 14, pp. 972–980, May 2006.

[8] S. R. M. Prasanna, D. Govind, K. S. Rao, and B. Yenanarayana, "Fast prosody modification using instants of significant excitation," in *Proc Speech Prosody*, May 2010.

[9] K. S. R. Murty and B. Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. Audio, Speech and Language Process.*, vol. 16, no. 8, pp. 1602–1614, Nov. 2008.