An acoustic exploration of event construals in Bengali language

S Chowdhury\textsuperscript{1*}, P K Gayen\textsuperscript{1}, U Sarkar\textsuperscript{2}, A Banerjee\textsuperscript{2,3}, S Sanyal\textsuperscript{1,3}, S Karmakar\textsuperscript{1} and D Ghosh\textsuperscript{2}

\textsuperscript{1}School of Languages and Linguistics, Jadavpur University, Kolkata: 700032.
\textsuperscript{2}Sir C.V. Raman Centre for Physics and Music, Jadavpur University, Kolkata: 700032
\textsuperscript{3}Rekhi Centre of Excellence for the Science of Happiness, IIT Kharagpur, West Bengal, 721301

*Email: spandan.gyan@gmail.com

Abstract. The world is full of events, and events are construed. One of the major research questions, therefore, seeks to understand the way events are construed through language. Event construals involve the syntacto-semantic properties of certain linguistic categories such as Verb, Tense, Aspect, Modality, etc. Serial Verb Constructions (SVC) and Complex Predicate Constructions (CPC) are no exception to this very fact. In this work, we look forward to compare both linear and non linear acoustical features generated from SVC and CPC events found in sentences of Bengali language. For this, we recorded 60 common utterances of Bengali language containing SVC and CPC events individually (around 36 of them belong to SV category while the rest belong to CP category), from 1 male and 1 female native Bengali speakers. The serial verb construction in these utterances may contain two (2) serial verbs in either simultaneous or sequential order, while few sentences are such that they can be interpreted both as SV or CP events. The main objective is to look for robust acoustic features which lead to perceptual categorization of events as SV or CP with a particular linguistic background. Various linear features like MFCC (Mel frequency Cepstral Coefficients), spectral skewness/ energy, pause duration, pitch profile and nonlinear features like Fractal Dimension (FD) have been employed for the classification purpose. This work is a pilot study of an ongoing project which looks to explore the concept integrating capacity of human brain in terms of Syntactic Compositionality or Semantic Combinatorics in a complex sentence. This preliminary acoustic study reveals interesting new results in terms of perceptual linguistic representation of the event construals.

Keywords: Event Construals, SVC, CPC, Bengali language, Acoustic Analysis, MFCC, Hurst Exponent, Serial Verbs, Complex Predicates

1. Introduction

The double verb construction has always been a matter of controversy in the field of linguistics when we consider the case of South-Asian languages - i.e., in which domain they are formed, in syntax or in morphology. Also, the status of these constructs, i.e., whether they should be considered as Phrases (or Clauses), or are they Words, has also been a matter of debate. The South Asian languages belong to four broad families namely; Indo-Aryan, Dravidian, Tibeto-Burman and Austro-Asiatic. Inspite of certain commonality in the syntactic-semantic features, all the languages have their own individualities [1, 2]. The Indo-Aryan and Dravidian languages in context to double verb construction which have been extensively studied in literature are Hindi-Urdu [3-6], Bengali [7-9], Punjabi [10, 11] and Marathi [12]. In this work, we intend to study two specific linguistic representations of Bengali language - Serial Verb Construction (SVC) and Complex Predicate Construction (CPC) in both acoustical and neuro-psychological paradigm.
Contemporary research on events can be conceived from four different (though interrelated) dimensions: First one of these is heavily influenced by the analytical rigour of the philosophical logic, focusing mainly on the literal significance of the symbols representing events, as is the work of Vendler [13]. In contrast to Vendler’s concern of identifying the criteria for defining different event types, Davidson [14] was primarily concerned about the way event is individuated. In recent times, this tradition has resulted into the investigation of sub-eventual intricacies [15]. Along the second dimension, researchers are interested in investigating how intra-sentential realizations of events are useful in constructing larger chunks of events at the level of discourse. In this tradition we can cite the work of van Lambalgen and Hamm [16]. Third dimension of research on event semantics asks questions of following types: How do people identify the temporal parts of events and the relations amongst those parts? Research has shown that observers spontaneously encode activity in terms of their temporal structure and subparts, and that this structure influences how they talk about activity and how they remember it later. A comprehensive overview along this line can be found in Zacks and Tversky [17].

On the acoustic analysis of specific linguistic features, Boersma [18] states that “Acoustic analysis - once a method used primarily within the domain of phonetics, has become an increasingly necessary skill across the field of linguistics". While the phonologists require acoustic data to substantiate their theoretical arguments, sociolinguists tend to characterize vowel shifts and mergers in terms of their acoustic properties, and psycholinguists draw on acoustic analysis techniques for construction of input stimuli during their experiments, neurolinguists on the other hand use standardized acoustic signals as input stimuli to gain insights into how the linguistic parameters are comprehend/processed in human brain. Till date, most research on the application of acoustics in linguistics research have been limited to the extraction of several low-level features from acoustic waveforms - such as pitch contour, intensity mapping and analysis of formants. These have mostly been used to characterize the vowels, duration of utterances, and also acoustic properties of speech, such as periodicity, intensity, and spectral qualities. On the other hand, acoustic feature analysis from speech signals have gone a long distance using long-range features such as lexical, prosodic, and discourse related habits [19-22]. Prosodic features include the rhythmic and intonational properties in speech, as examples are voice fundamental frequency (F0), F0 gradient (pitch), intensity (energy) and duration while features like Mel Frequency Cepstral Coefficients (MFCC) as a filter bank approach have long been used in speech recognition tasks. In this work, for the first time we have used several acoustic spectral features such as spectral tilt, spectral centroid, pause duration between the two Serial verbs (SVs), MFCCs etc. to classify and characterize the acoustic waveforms generated from the event construals.

Fractal analysis of audio signals was first performed by Voss and Clarke [23], who analyzed amplitude spectra of audio signals to find out a characteristic frequency $f_c$, which separates white noise (which is a measure of flatness of the power spectrum) at frequencies much lower than $f_c$, from very correlated behavior ($\sim 1/f^2$) at frequencies much higher than $f_c$. Any speech data is a quantitative record of variations of a particular quality over a period of time. It is a well-established experience that naturally evolving geometries and phenomena are rarely characterized by a single scaling ratio; different parts of a system may be scaling differently. That is, the clustering pattern is not uniform over the whole system. Such a system is better characterized as ‘multifractal’ [24]. A multifractal can be loosely thought of as an interwoven set constructed from sub-sets with different local fractal dimensions. Real world systems are mostly multifractal in nature. Speech and music signals too, have this non-uniform property in its movement [25-27]. In this work, nonlinear analysis in the form of Multifractal Detrended Fluctuation Analysis (DFA) proposed by [28], have also been performed on the two types of utterances. In this way, we intend to classify the two event construals on the basis of high level acoustic features so that we can have an idea of their perceptual correlates before moving on to find out their neural correlates, which is the final outcome of the present project.
2. Methodology
2.1. Experimental Dataset:
The dataset for our analysis included a corpus of 60 sentences which have been categorized into Serial Verb Construction (SVC) and Complex Predicate Constructions (CPCs) based on the event construals. The recordings were taken from 1 Male and 1 Female participant, who were asked to record the sentences provided along with a few distractors, which ensured the respondents are not primed while recording the data. We recorded the data according to our framework using the standard Praat software tool. The acoustic waveform corresponding to the focus area from each of the sentences was extracted and annotated as follows:

![Waveform](image1)

**Figure 1.** Waveform corresponding to Focus verbs of (a) “Shyam notun godite shuye dekhlo” (a serial verb presented simultaneously) (Translation: Shyam lay down on the new mattress to test it) (b) “Ram Shyamke deke elo” (a serial verb presented sequentially) (Translation: Ram asked Shyam to come and then came back himself) and (c) “Riju board theke lekhata muchhe phello” (a complex predicate) (Translation: Riju erased the writing from the board)

2.2. Acoustic Spectral features

2.2.1. Pitch Contours: F0 contour reflects a proper temporal orientation of complex predicates and serial verbs in Bengali language. A window length of size 0.015s with hop size 0.005s is taken to extract F0 contour. Statistical features like, higher order moments (Skewness, kurtosis) and trend
analysis are done to extract out significant data to represent the verbs. We have extracted the pitch profile corresponding to the focus verbs in all the 60 target sentences using standard MATLAB chroma based features [30] and compared the pitch contour corresponding to the Serial verb and Complex predicate constructions.

2.2.2. Spectral Flatness: spectral flatness or tonal coefficient is ratio between geometric mean and arithmetic mean of power spectrum. It provides a measure of how tone-like an audio is. A high value of spectral flatness represents white noise characteristics of audio spectrum.

\[ \text{spectral flatness } F = \frac{\left(\prod_{n} F(n)\right)^{\frac{1}{N}}}{\frac{1}{N} \sum_{n} F(n)} \]  

(1)

2.2.3. Spectral Centroid: Spectral centroid quantifies the brightness of a spectrum and can be evaluated by extracting the center of gravity of a spectrum by means of its magnitude and frequency of a bin.

\[ \text{spectral centroid } C = \frac{\sum_{n} n F(n)}{\sum_{n} F(n)} \]  

(2)

2.2.4. Pause duration: The pause between two spoken words is considered to be a very important feature while evaluating the perceptual correlates of the spoken utterance. In this particular study, we have evaluated the inter-verbal pause duration for all the 60 utterances related to SVC and CPCs. The pause duration has been found to be minimal for CPCs while it is on the higher side when we consider the SVCs, while for simultaneous verbs in SVCs, it is again on the lower side.

2.3. Multifractal Detrended Fluctuation Analysis:

The time series data obtained from the sound signals are analyzed using MATLAB and for each step, an equivalent mathematical representation is given using the prescription of Kantelhardt et al [28]. In MFDFA technique, the whole length of the signal is then divided into \( N_s \) number of segments consisting of certain number of samples.

The local RMS variation for any sample size \( s \) is the function \( F(s,v) \). This function can be written as:

\[ F^2(s,v) = \frac{1}{s} \sum_{i=1}^{s} [Y[(v-1)s+i]-y_v(i)]^2 \]  

(3)

The q-order overall RMS variation for various scale sizes can be obtained by the use of following equation:

\[ F_q(s) = \left[ \frac{1}{N_s} \sum_{v=1}^{N_s} \left\{ F^2(s,v) \right\}^{\frac{q}{2}} \right]^{\frac{1}{q}} \]  

(4)

The scaling behaviour of the fluctuation function is obtained by drawing the log-log plot of \( F_q(s) \) vs. \( s \) for each value of \( q \).

\[ F_q(s) \sim s^{h(q)} \]  

(5)

The \( h(q) \) is called the generalized Hurst exponent. The presence or absence of long range correlation can be determined using Hurst exponent. A monofractal time series is represented by varying values of \( q \) with \( h(q) \) (Fig.2(a)). The singularity spectrum \( f(\alpha) \) is related to \( h(q) \) by

\[ \alpha = h(q) + q h'(q) \]

\[ f(\alpha) = q! |a - h(q)| + 1 \]

Where \( \alpha \) denoting the singularity strength and \( f(\alpha) \), the dimension of subset series that is characterized by \( \alpha \). The width of the multifractal spectrum essentially denotes the range of exponents. The spectra can be characterized quantitatively by fitting a quadratic function with the help of least square method in the neighbourhood of maximum \( \alpha_0 \),

\[ f(\alpha) = A(\alpha - \alpha_0)^2 + B(\alpha - \alpha_0) + C \]  

(6)

Here \( C \) is an additive constant, \( C = f(\alpha_0) = 1 \) and \( B \) is a measure of asymmetry of the spectrum. So obviously, it is zero for a perfectly symmetric spectrum. We can obtain the width of the spectrum very easily by extrapolating the fitted quadratic curve to zero.
Width $W$ is defined as,

$$W = \alpha_1 - \alpha_2$$

with

$$f(\alpha_1) = f(\alpha_2) = 0$$

(7)

The width of the spectrum gives a measure of the multifractality of the spectrum (Fig. 2(b)). Greater is the value of the width $W$, greater will be the multifractality of the spectrum. For a monofractal time series, the width will be zero as $h(q)$ is independent of $q$.

Figure 2 (a) and (b) represent the variation of $h(q)$ with $q$, and $f(\alpha)$ with $\alpha$ respectively for a sample utterance taken. The multifractal spectral width or the acoustic complexity in each case is computed from the width of the multifractal spectrum shown in figure 2 (b).

3. Results and Discussion

The pitch contour corresponding to the 36 SVCs (out of 60) have been plotted in figure 3(a), while the pitch contour corresponding to the rest 24 CPCs have been plotted in figure 3(b).

Figure 3. Pitch contour corresponding to (a) 36 SVCs and (b) 24 CPCs respectively
The pitch profile corresponding to the two types of event construals shows clear distinction in the way they are acoustically perceived in the human mind. While for the complex predicates, there is a single peak in the pitch contour, the presence of double peaks in the pitch contour of serial verb constructions indicate the processing of these verbs differently. In case of complex predicates, the presence of a discrete high pitched sound in the end is consistently visible, which is missing in case of serial verb constructions. The two distinct peaks in case of SVCs imply the differential processing of the two verbs individually in the human perceptual system. In this way, using simply a pitch profile analysis, distinction of the two event construals has been done.

By method of statistical analysis, we obtained an 8-dimensional feature space of each pronounced verb instance using the different acoustical features detailed in the Methodology section above. We applied PCA on this feature space to point out the 3 prime Eigen bases with maximum Eigen values. Figure 4. shows the instances plotted on this reduced 3-dimensional space.

![3D Acoustic Feature plot corresponding to the 60 utterances of event construals](image)

A very interesting clustering pattern is observed in the 3D acoustic feature plot of the different event construals put to test in this work. While the values of the Serial Verbs are widely distributed across the feature plot, the values corresponding to the Complex Predicates are somewhat clustered in different regions. Also, a number of overlaps or ambiguous values are noticed in these two cases implying the feature extraction techniques employed are not able to exclusively categorize the two events in these cases. The widely scattered values of SVCs denote the perceptual variance as observed in the different serial verb construals present in Bengali language while the clustering pattern in the CPCs denote the uniqueness of the two verbs representing a single event. Another interesting observation from the above plot is that in most cases the values of the simultaneous verbs in SVCs overlap with those of CPCs indicating the acoustical similarity that is present in the utterances of these two forms. Also, the values of simultaneous verbs in the feature space are not that spread as compared to that of their sequential counterparts. This indicates that acoustically speaking; the SVCs which are uttered simultaneously are somewhat similar in feature with that of the verbs uttered in case of CPCs in Bengali spoken language.
Next, for the nonlinear acoustic analysis, the multifractal width corresponding to the 60 utterances have been computed and reported in table 1 below. In this case, however, we find a third category from among the SVCs i.e. they can be categorized in simultaneous and sequential verb formations.

Table 1. Multifractal width corresponding to the focus verbs of 60 recorded utterances

| Complex Predicate | Serial Verb (Sequence) | Serial Verb (Simultaneous) |
|-------------------|------------------------|---------------------------|
| 1.72              | 1.43                   | 1.78                      |
| 1.53              | 1.73                   | 2.69                      |
| 1.56              | 1.45                   | 1.89                      |
| 1.80              | 1.62                   |                           |
| 1.65              | 1.34                   |                           |
| 1.23              | 1.50                   |                           |
| 1.06              | 1.48                   | 2.05                      |
| 1.65              | 1.50                   |                           |
| 1.56              | 1.19                   |                           |
| 1.56              | 1.53                   | 1.77                      |
| 1.60              | 1.30                   | 2.06                      |
| 1.57              | 1.46                   | 1.85                      |
| 1.46              | 2.09                   |                           |
| 1.44              | 1.16                   |                           |
| 1.31              | 1.46                   |                           |
| 1.34              | 1.46                   |                           |
| 1.39              | 1.40                   | 1.63                      |
| 1.55              | 1.62                   | 1.54                      |
| 1.50              | 1.61                   |                           |
| 1.45              | 1.32                   |                           |
| 1.92              | 1.43                   |                           |
| 1.53              | 1.57                   | 1.62                      |
| 1.70              | 1.60                   | 1.75                      |
| 1.75              | 1.42                   | 1.01                      |

The multifractal widths corresponding to the three event construals found has been plotted graphically in Fig. 5.

![Figure 5. Distribution of acoustical multifractal widths of the event construals presented](image)
From Fig. 5, it is evident that the acoustic complexity of the complex predicates, in general, is higher than that of serial verb constructions. Also, the SVCs which are simultaneously ordered report the highest value of acoustical signal complexity among the three groups found. In the few cases of simultaneous processing of the verbs, the acoustic complexity seems to be overlapping with that of CPCs. This is an indication that the event construals in these cases are not significantly resolved, but there exists a fuzzy area in which the brain processes ambiguously these two types of event construals. In this way, with the help of different linear and nonlinear acoustic signal processing methods, we have tried to quantitatively categorize the two most basic types of event construals that are present in the Bengali language.

4. Conclusion

The presence of serial verbs and complex predicates in South Asian language groups is well documented in literature. In case of SVCs, more than one verb is used to predicate more than one event; whereas in case of CPCs, two or more verbs are used to predicate one single event. In this paper, we look to find the acoustic correlates of these event construals using different linear and nonlinear signal processing techniques. The study gives the following interesting conclusions:

1. Pitch profile analysis of the CPCs show a single concentrated peak, signifying the processing of the two verbs as a conjugated whole. The SVCs present two distinct peaks in the pitch contour with a definite pause in between, signifying the processing of two consecutive verbs as separate linguistic entity.

2. The different acoustic features put together in a 3D feature plot demonstrate specific clusters for the three different forms of event construals in question. While the SVC features are widely distributed over the entire dimension of the plot, the CPCs form clusters in different regions showing their unique identity of treating the two verbs as a single event. The simultaneous verbs in SVCs, in most cases, lie acoustically very close to that of CPCs, indicating that feature-wise they are very similar, although they are listed as separate event construals in the literature.

3. The nonlinear analysis in the form of multifractal spectral analysis reports the presence of three distinct classes of event construals whose acoustic complexities are statistically different from one another. Here, we see that the SVCs in which verbs encode simultaneous events have the highest acoustic complexity and in few cases, specific overlaps with that of CPCs are found. In this way, we have been able to categorize acoustically the two basic forms of utterances found in spoken Bengali language. It would be interesting to corroborate these results with that of a human response and a neuro-cognitive study where the participants are made to listen to a pre-recorded set of SVCs and CPCs along with other distractor sentences. The EEG-ERP analysis will give us new insights into how these event construals are linguistically processed in the human brain. This pilot study is a precursor in that direction.

5. Acknowledgements

SC, SS and SK acknowledge DST CSRI, Govt of India for providing the funds related to this Major Research Project (DST/CSRI/2018/78 (G)). One of the authors, AB acknowledges the Department of Science and Technology (DST), Govt. of India for providing (SR/CSRI/PDF-34/2018) the DST CSRI Post Doctoral Fellowship to pursue this research work. SS further acknowledges the Acoustical Society of America (ASA) for providing the International Students Grant.

6. References

[1] Masica, Collin, P. 1976. Defining a Linguistic Area. Chicago. University of Chicago Press.
[2] Masica, Collin, P. 1991. The Indo Aryan Languages. Cambridge. CUP.
[3] Nespoli, Helmut. 1997. Lokahhaaratii: Hindii kriyaa-kosha. Lokbharti Prakashan, Allahabad.3 26
[4] Butt, M. 1997. Complex Predicates in Urdu, In Alsina, A., J. Bresnan and P. Sells (eds) Complex Predicates, Stanford. CSLI Publications.
[5] Arora, H. 1979. Aspects of Compound Verbs in Hindi. M.lit dissertation, Delhi University.
[6] Hook, P. 1974. The Compound Verbs in Hindi. The Michigan Series in South and Southeast Asian Language and Linguistics. The University of Michigan
[7] Singh, U.N. 1998. Compound verbs in Selected South Asian Languages. In Khokhlova, L.V. and Atul
[8] Ramchand, G. 1990. Complex predicate formation in Bangla. In proceeding of the Ninth Annual West Coast Conference on Formal Linguistics.
[9] Dasgupta, P. 1977. The internal grammar of compound verbs in Bangla. Indian Linguistics, 38:3, 68-85.
[10] Akhtar, R. N. 2000. Aspectual Complex Predicates in Punjabi. Ph.D Dissertation. University of Essex, Colchester.
[11] Bhatia, T. K. 1993. Punjabi: A cognitive-descriptive grammar. Routledge. London and New York.
[12] Pandharipande, R. 1990. Serial Verb Construction in Marathi. In A. Zwicky and B. Joseph (eds.), When Verbs Collide: Papers from the 1990 Ohio State Miniconference on Serial Verbs, 178-199.
[13] Vendler, Z. (1957). Verbs and times. The Philosophical Review, LXVI:143-160.
[14] Davidson, D. (1967). The Logical Form of Action Sentences. Oxford University Press.
[15] Verkuyl, H. J. (1993). A Theory of Aspectuality. Cambridge University Press, Cambridge.
[16] van Lambalgen, M. and Hamm, F. (2005). The Proper Treatment of Events. Blackwell Publishing, Oxford.
[17] Zacks, J. M. and Tversky, B. (2001). Event Structure in Perception and Conception. Psychological Bulletin, 127(1):3-21.
[18] Boersma, P. (2014). 17 Acoustic analysis. Research methods in linguistics, 375.
[19] Shriberg, E, L Ferrer, S Kajarekar, A Venkataraman, and A Stolcke. "Modeling prosodic feature sequences for speaker recognition." 46 (2005): 455–472. Print.
[20] Sen, Nirmalya, T.K Basu, and Hemant. A Patil. "New Features Extracted from Nyquist Filter Bank for Text Independent Speaker Identification." Annual IEEE India Conference (INDICON). 978-1-4244-9074-5/10. (2010): 1-5. Print.
[21] W. M. Ng, Raymond, tan Lee, Cheung Chi Leung, Bin Ma, and Haizhou Li." Analysis and Selection of Prosodic Features for Language Identification. 978-0-7695-3904-1/09. (2009): 123-128.
[22] Kockmann, Marcel, Lukas BurgetLast, and Jan Honza Cernocky. "INVESTIGATIONS INTO PROSODIC SYLLABLE CONTOUR FEATURES FOR SPEAKER RECOGNITION." ICASSP 2010. 978-1-4244-4296-6/10.2010 (2010): 4418-4421. Print
[23] Voss, R. F., & Clarke, J. (1975). 1/f noise in music and speech. Nature, 258(5533), 317-318.
[24] Lopes, R., & Betrouni, N. (2009). Fractal and multifractal analysis: a review. Medical image analysis, 13(4), 634-649.
[25] Sanyal, S., Banerjee, A., Patranabis, A., Banerjee, K., Sengupta, R., & Ghosh, D. (2016). A study on Improvisation in a Musical performance using Multifractal Detrended Cross Correlation Analysis. Physica A: Statistical Mechanics and its Applications, 462, 67-83.
[26] Sanyal, S., Banerjee, A., Nag, S., Sarkar, U., Roy, S., Sengupta, R., & Ghosh, D. (2020). Tagore and neuroscience: A non-linear multifractal study to encapsulate the evolution of Tagore songs over a century. Entertainment Computing, 100367.
[27] Sanyal, S., Nag, S., Banerjee, A., Sengupta, R., & Ghosh, D. (2019). Music of brain and music on brain: a novel EEG sonification approach. Cognitive neurodynamics, 13(1), 13-31.
[28] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. Physica A: Statistical Mechanics and its Applications, 316(1-4), 87-114.
[29] F. Martinez, A. Guillamon and J.J. Martinez, Vowel and Consonant characterization Using Fractal Dimension in Natural Speech. Non-Linear Speech Processing (NOLISP 03), 2003
[30] Müller, M., & Ewert, S. (2011). Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), 2011. hal-00727791, version 2-22 Oct 2012.*