 CHAPTER 1
INTRODUCTION

Words are presented as syllable sequences in speech. However, speech perception is said to be phonemic in nature. In continuous speech too, the phonemes are 'coarticulated', e.g. /k/ and /o/ are coarticulated during the articulation of the word [komol] (‘soft’) in Bangla (B.) or [komol] (‘soft’) in Hindi (H.). These coarticulated phonemes represent higher structures i.e., syllables as a part of explanatory adequacy in linguistics. This dogma, attached to models of spoken language perception and analysis, is extended to studies of cognitive neuroscience of language as well. As a result, there are competing theories about whether phonemes act as the default mode for perception and recognition of words (as is assumed), or whether it is the syllable that forms the perceptual core.

We perceive concatenated phonemes as perceptual complexes, e.g. /k/ and /o/ are perceived as /ko/. Does this mean coarticulated phonemes have valid acoustic signatures or neural correlates explaining the binding of two phonemes? Further, in spoken word conditions, it is easier to recognise a word when presented as syllables. Does this mean that the acoustic signatures of syllables are stored as word-internal properties? The present perception theories are unable to answer these questions as perception and recognition are proposed to be two separate processes in lexical access.

This thesis investigates the perceptual domain of phoneme and syllable representation in terms of spectral and temporal parameters that can be tested for binding across phonetic, phonological and neural levels of analysis. In addition, it also looks into whether syllables are word-internal properties or not in the phonological code.

1.1 Syllable as a unit of analysis in linguistics, psychology and neurobiology

Syllable is one of the most researched problems in psycholinguistics, linguistics and currently in cognitive sciences. However, the primary issue has remained elusive: how is the syllable represented in the brain? In other words, how do we recognize the syllable – as a chunk, or as a progressively incremental unit of phoneme sets? The current
literature clearly has distinct standpoints depending on what is getting encoded in the brain, phoneme-centric perspective (Marslen-Wilson 1989, Pisoni & Luce 1987), syllable centric perspective (Mehler 1981, Greenberg 2003, Oudeyer 2002, Poeppel 2005). The phoneme-centric perspective considers phoneme as the unit of primary encoding while syllable-centric perspective considers syllable as the unit of primary encoding in the hierarchy of speech processing.

As a linguistic unit, syllable has enjoyed a special status in both phonetic and phonological analysis. Although replete with controversies, it is largely accepted as a unit of phonological analysis. The existence of syllable as a unit of analysis was overlooked in the post-Chomskian contribution [Sound Patterns of English (SPE) Model, Chomsky & Halle 1968]. It was reintroduced in the 1970s (Hooper, Vennemann, etc.) as a domain of application of the phonological rules and also as a unit of phonotactic restrictions. However, its status is debated as a physical unit in the phonetic analysis. This has often been primarily attributed to the variable nature of speech and the astonishing ability of co-articulating any given sequence of consonants and vowels continuously. The confusion of classifying it as a separate hierarchical organization has increased owing to a high variability in duration, length, and complexity. Attributes such as sonority have also lent to further confusion in the absence of appropriate quantification parameters.

As a psycholinguistic unit, the syllable has been linked to speech perception and processing in the mental lexicon. The ability for recognition and production has been the primary tools of investigating syllable. Consequently, the innate ability to recognize and articulate has been linked to both short term and long-term memory on one hand and to access words in the lexicon on the other. Clearly, there are two different levels of understanding speech processing and hence syllable processing. The brain constitutes the physical level where neural activation patterns and response times constitute investigation criteria and the abstract level involves the mental lexicon where the constituents of language are stored.

Neural representation for language assumes one to one mapping with the structural and functional regularities of speech. There has been a long standing debate of whether the speech and language abilities emerged from a ‘language module’ unique to human brain (Wilkwins & Wakefeild1995, Liberman & Mattingley 1985) or through the
elaboration of more basic sensory, motor and cognitive neural mechanisms common to human and non-human species (Fitch, Miller & Tallal 1993).

Neurological and behavioural studies on individuals with language-impaired populations have provided basis for structural demarcation of the regions responsible for both language and higher order cognitive processes. Evidence has shown concurrent deficits in processing rapidly changing acoustic cues and speech following left hemisphere damage (Tallal & Newcombe 1978) suggesting a direct association between perception of acoustic cues and speech perception.

Studies on language learning-impaired (LLI) children show that an impaired auditory cortex fails to process rapid auditory processing tasks even when non-linguistic stimuli are used (Tallal & Piercy 1973, Fitch, Miller & Tallal 1993). Recent studies (Poeppel 2003) propose that speech comprehension is correlated with temporal response patterns in the auditory cortex. Since comprehension of the speech depends on integrity, it provides a logical basis for argue that the temporal envelope contains information that is crucial for identification of phonemes, syllables, words and sentences. However, it carefully evades the nature of spectral content i.e., whether the temporal envelope contains acoustic information as a grouping of distinctive features leading to perceptual binding or integration of distinctive features leading to binding of the spectral cues in the temporal envelop.

Apart from this, Scott and Wise (2003 a, b) and Scott and Johnsrude (2003), have outlined evidence for hierarchical speech processing in human beings. According to them, hierarchical perceptual processing is a characteristic of primate perceptual cortex, and a progression of processing can be seen, from registration of signal in primary auditory cortex, to amplitude modulations and spectral structure in lateral auditory areas to more speech-like spectral and frequency modulations in yet more lateral regions. Neuroimaging experiments reporting activation of similar areas (planum temporale, superior temporal gyrus, etc.) both in response to syllables (Chen, Chen & Dell 2004), contrastive features, place of articulation (Obleser, Lahiri & Eulitz 2004) and frequency modulations (Zattore 2000, 2007) add to the debate of recognizing syllable as a physical or analytical unit of organisation.

Consequently, the argument for whether a syllable as a unit stored in the 'mental syllabary' as an independent interactive module capable of processing of speech online
(Levelt 1992) or as an added correlate in the internal structure of word (Chen, Chen & Dell 2004) becomes more pertinent. Neural representations of speech and speech-like processes have been progressively plotted in the primary auditory cortex, superior temporal sulcus (STS) and superior temporal gyrus (STG) lateral to primary auditory cortex providing us with evidence of anterior and posterior streams of information processing in the brain. Given these conditions of parallel hierarchical processing any recognized stimulus requires both temporal and spectral resolution as well as integration.

Syllable is no less complex when it comes to its dynamical state, acoustic variability and efficient access from memory, allowing higher level spatial, temporal and perceptual correlations to produce real-time input and output. The rich variation of syllable types found in human languages also contributes towards having a broad definition of syllable as a theoretical necessity, and often this generality contributes to ambiguity in understanding syllable as a discrete unit in speech processing. The syllable can assume structures V, VC, CV, CVC, VCV, etc. This brings out the polymorphic character of syllabic segmentation. Such richness only brings us to another set of questions- why do we need so many types of syllables at production and recognition levels after all? Do we have a canonical syllabic form? How are they neurally represented? Can such polymorphism be favoured as the basis of a unit of storage too?

Therefore the primary focus of this study is to find answers for the question- are complex structures such as syllable represented in the brain? More specifically, this addresses two questions: firstly, Are CV and VC represented similarly? And secondly, do CV and CVC have representations in the same tier, i.e., do they have same level of complexity?

1.2 Justification for mapping complex structures like the syllable

Analysis of syllable structure (i.e. having onset, nucleus and coda) in many languages shows that the notion of syllable allows various deviations of the basic scheme. Traditional descriptive and theoretical terminology describes this as ‘complex onset’. Though this complexity is variously accounted for by the sonority hierarchy theory, it is not clear how they are represented in the brain. The reason for such ambiguity in analysis may be attributed to the number of theories explaining syllabification and the syllable boundary placement. The following table provides an example for the same. As a result
syllabification has been understood as a surface phenomenon. But the syllabic has not been investigated with respect to adaptational-functional parameters.

Table 1.1 Syllabification patterns

<table>
<thead>
<tr>
<th>Syllabification model</th>
<th>Syllabification in Hindi</th>
<th>Syllabification in Bangla</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal onset principle</td>
<td>/ok.kɔr/</td>
<td>/ok.kʰɔr/ (or) /ok.kʰɔr/</td>
</tr>
<tr>
<td>Core syllabification model</td>
<td>/ə.kʃɔr/</td>
<td>/o.kkʰɔr/ (or) /o.kkʰɔr/</td>
</tr>
<tr>
<td>Maximum open syllabicity</td>
<td>/ək.ʃɔr/</td>
<td>/o.kkʰɔr/ (or) /o.kkʰɔr/</td>
</tr>
<tr>
<td>Sonority syllable model</td>
<td>/ək.kɔr/</td>
<td>/ok.kʰɔr/ (or) /ok.kʰɔr/</td>
</tr>
</tbody>
</table>

Further, the syllable has always been regarded as a functional property in linguistic analysis in order to account for morphological problems that cannot be explained otherwise and a domain of application of phonological rules. But in order to account for the actual variations in structure in auditory processing, a form for representation of complex structures both for psycholinguistic and neurobiological bases become a necessity.

The phonemic models of spoken language do not provide an accurate characterization of the spoken language especially when viewed as a unit of recognition. This is evident from ASR (Automated Speech Recognition) models based on phonemic models of speech perception as they have lot of difficulty in decoding the speech. Further continuous annotation and analysis of five hours of SWITCHBOARD corpus shows that the speakers tend to ‘delete’ segments in the pronunciation of many words (see Greenberg 2002 for a review). Phonemic realizations also differ from the canonical representation (standard variety) of the phoneme in natural speech as a result of individual speaker (context dependent) and across speaker variability due to linguistic background, accent, speaking rate, acoustic and semantic environments.

Empirical studies such as acoustic-phonetic study of English language corpus (Greenberg 2004), Hungarian (Basbóll 2005), Polish, Dutch, etc. have also indicated that there are complex surface representations that cannot be explained by the prevalent definitions of syllable, e.g. onset clusters at intervocalic positions found in these
languages. Onset clusters at intervocalic positions are also found in Indo-Aryan languages, Bangla and Hindi. There surface representation though has clusters but the nature of the deep structure is not known. Although the theory of having empty nuclei (the deeper structure nuclei are not expressed in the surface form), due to phonotactic consideration appears to be an interesting extension from Polish and Dutch, not much research has been carried out about the structure of deeper representation. This raises the question, ‘Are deep structure representations also syllabic?’

Syllable as a unit of perception owing to its higher degree of stability (Greenberg et al. 2002) has an advantage over the phoneme models of perception. The production models (Levelt, Roelofs & Meyer 1999, Cholin et al. 2006, Guenther et al. 1999) also favour the motor programs for the syllables (stored in mental syllabary) as a basis of speech production over phonemes (although the processing of the spoken words during lexical access is strictly phonemic). Further, according to Greenberg et al. (2002), the syllable is able to absorb some ‘extraphonetic properties pertinent to the prosody of the utterance’. Further, syllable offers a dual level of characterization in terms of decomposition to phonemes and relative ‘prominence’ of syllables in utterance. This allows phonemic distinction and prosodic implementation (accent).

The perceptual paradigms have also addressed the problem of phoneme and syllable representation they have basically used the syllabic conditions where the listener is carrying out phoneme recognition tasks during contrasting conditions e.g., the perception of phonemes is largely based on the distinction of place and manner of articulation during production of consonants which agrees to the basic criteria of minimal alternation and maximal contrast in representation. Hence it is imperative to look at the syllable as a natural unit of complex representation of language. However owing to the fact that the present neurobiological paradigms commonly agree to the phoneme representation as the minimal unit that is represented as the ‘Central Sound Representation System’ the hypotheses for this thesis assume phonemes to be the basis of perception and are capable of syllable and word representation and disproves them.

1.3 The Nature of data

Evidence for posting syllabification and syllable organization comes from the following domains-Phonological generalization, Particular facts of speech production and
speech perception, Subject judgements about speech and their ability to recognize and decomposes words into sub-morphemic parts. However, in order to evaluate the presence or absence of any form in the mental representation, it is imperative to look for parameters that can serve as quantitative marker in both the phonological and the phonetic domain so that they can be grounded in the cognitive system.

The present thesis evaluates the phonological generalization of syllable with the acoustic-phonetic data in Bangla and Hindi with respect to binding parameters critical for perception of syllable as a word-internal property. Subsequently the binding parameters were used to look for syllabic representation using 'reseriation' (defined as the tendency to interchange the sequence of the syllables in a bisyllabic condition in fast speech with reduced onset and offset effects) as a tool to study the representation of the syllable in the brain.

The audio samples were recorded in the acoustically treated room with a condenser microphone in a DAT recorder (Sony) and digitised at 44100 Hz and transferred to the computer with an external soundcard (Sound Blaster, Creative, USA). The samples were downsampled to 11025 Hz, normalized for amplitude at 40 dB with Adobe Audition 2.0 (Adobe Systems, USA) and Audacity. The preliminary audio samples were annotated and segmented in PRAAT and the formants of the vowels analysed.

Subsequently, the audio samples were subjected to wavelet transform, which resolves and analyses the audio samples in a multiresolution time window. Features extracted from the acoustic analysis of spoken syllables and words were used for recognition tests for investigating syllable representations. EEG has also been used as a tool to study the change in electronegativity resulting due to the auditory presentation of the sound samples. EEG provides excellent temporal resolution and serves to be a better tool to investigate syllable processing.

1.4 Hypotheses

The present thesis investigated the claim that syllable is composed of more than one dissimilar phonemes and that phonemes act as the central auditory sound representations in spoken word comprehension. The investigation centred around two main hypotheses:
i. Phonemes have independent identity at the level of auditory object analysis and hence are singular percepts with featural and temporal integration.

ii. Phonemes are highly localized in the short-term and long-term memory traces during lexical activation and spoken word recognition and comprehension.

1.5 Outline of the thesis

The present thesis has six chapters. Chapter 1 provides a general overview about the topic and purpose of present study. Chapter 2 gives a brief overview of the available definitions of syllable, principles and models of syllabification. Further, it describes and evaluates the available perception and neurobiological research in the light of linguistic notion of syllable. The meta-analysis provides a basis for stating the hypothesis. Chapter 3 evaluates the first hypothesis (listed in the Section 1.4) and provides acoustic, behavioural and temporal information about the binding parameter of syllable (and hence disproves independent existence of consonantal phonemes at onset and coda positions). Chapter 4 evaluates the second hypothesis (listed in the Section 1.4) and provides evidence that syllable structures are word internal properties. Chapter 5 evaluates the conceptual, linguistic and biological and psychological basis of binding parameters of syllable (EDD) and argues that syllable is a better candidate for being an auditory object based on the constraints derived from the acoustic analysis and behavioural studies. It proposes syllable as a neuro-cognitive biomarker and presents the minimal representation of syllable in the model. Chapter 6 concludes the thesis with the main results, constraints, implications, reservations and future research in syllable cognition. The conclusion has three inferences:

a. The syllable structure has a complex representation in the perceptual domain, where articulatory ‘release’ is represented by the Energy Density Decay profile. This is evident in both actual and low pass filtered speech samples.

b. Two different levels of representation are evident from the type of EDD slope i.e., CV and VC have only one EDD slope (although the types of EDD slopes are different) while CVC and VCV have two EDD slopes (the order of the EDD slope type varies between them). CV has a positive attenuation of energy where as CVC has both positive (release of onset in nucleus) and negative attenuation of energy
(i.e., the energy is accumulated before a silent release). This negative attenuation of energy can be used as a tool to explain resyllabification in continuous speech.

c. Syllabicity is word-internal as evident from reseriation to access two different semantic representations.

The findings show an adaptationist-functionist approach to syllabification. It proposes a generic syllable (critical duration and release) that may be extended as a natural unit of analysis. It is closer to the concept of ‘akshara’ in Indian linguistic tradition. Further, such a constraint-based representation that is grounded in acoustic-phonetic parameters can provide insights about its representation in the brain.