To understand the meaning of a natural language, the system must have the linguistic knowledge of that language. A language understanding programme having adequate knowledge about the structure of the language i.e. words, phrases and sentences, must know the meaning of the sentence and the context in which they are being used. For understanding the natural language the requirement of linguistic knowledge are classified as follows:

**Syllable:** Syllables are considered the phonological building block of words. They can influence the rhythm of a language, its prosody and stress etc. For example, the word *water* is composed of two syllable *wa* and *ter*.

**Morpheme:** Morpheme is smallest linguistic structure which bears meaning. The English word *unbearable* can be analyzed as comprising three morphemes, the prefix *un*, the suffix *able* and the root *bear*. The root is a free morpheme, because it may occur alone. The first two are bound morphemes, because they do not occur alone, but are always attached to other morphemes. In spoken language, morphemes are composed of *phonemes* (the smallest linguistically distinctive units of sound), and in written language morphemes are composed of *graphemes* (the smallest units of written language).

**Phoneme:** Phoneme is the smallest contrastive unit in the sound system of a language. A distinction between meaning of words, pronounced in one or more ways depending on the numbers of units called allophones. For example, the
phoneme /r/ and /l/ serve to distinguish the word *rip* from the word *lip*. Here phone [r] and [l] are contrast in identical environments.

**Allophones:** An allophone is a phonetic variant of phonemes in a particular language. In English i) [p] and [pʰ] are the allophones of the phoneme /p/ ii) [t] and [tʰ] are the allophones of the phoneme /t/.

**Word:** Word is the smallest free form related to the language with semantic and pragmatic content. Word may refer to a spoken word or a written word, or sometimes, the abstract concept behind either. Spoken words are made up of phonemes, and written words of grapheme.

**Semantic:** Semantic is the study of meaning in a language. This knowledge is concerned with the meaning of words and the phrases and how they combine to form meaning of a sentence.

**Pragmatic:** Pragmatic is the ways in which context contributes to meaning. The knowledge relates the use of a sentence in different contexts. It also signifies how the contexts affect the meaning of the sentence.

**Syntactic:** The syntactic knowledge relates or defines how words are put together or structured to form grammatically correct sentence in a language.

**Phonology:** Phonology is the study as to how sounds functions in a language.

**Phonetics:** Phonetics is the study of the sounds of speech as physical events.
Phonotactics: Phonotactics is the study of the permitted sequences of phonemes within the syllable in a particular language.

Morphological: It is a lexical knowledge that relates to the word construction from basic unit called morphemes, which is the smallest unit of meaning.

1.2 LITERATURE REVIEW OF SPEECH RESEARCH

Speech perception and recognition is the process by which the sounds of speech are heard, interpreted and understood. The study of speech perception is closely linked to the fields of phonetics and phonology in linguistic. Research in speech processing seeks to understand how human listeners recognize speech sounds from the environment and use this information to understand spoken language. The results of these researches have been used to explore more possibilities in building machine systems that would provide utmost benefit to human kind.

During the first half of the 20th century, Fletcher (1922) and others at Bell Laboratories documented the relationship between a given speech spectrum and its sound characteristics as well as its intelligibility, as perceived by a human listener. Miller et al. (1951) had made a classic study on the intelligibility of English monosyllables as a function of the size of test vocabulary and degree of background noise. He shows that when the number of possible responses is fairly small, intelligibility is good, even at high relative levels of background noise. But when the response set is large say all the monosyllables are in English; listeners are only
about 60% correct even in quite good listening conditions. Pickett & Pollack (1963) took a different approach to speech intelligibility. They showed that a segment from connected speech must be at least 800 ms long to be fully intelligible. Normally this represents at least two or three syllables. However, an experiment by Ladefoged & Broadbent (1957) had made an attempt to isolate particular influencing factors such as formant structure, that makes the distinction between two words (say) *bit* and *bet*, found that listeners judge the identity of the word mostly on their inherent formant structure.

The ability to distinguish a particular sound from other sounds happening simultaneously in the same environment, enables us to talk in a noisy place. The topic has been called the 'cocktail-party effect' by Cherry (1953). His experiments regarding attention reveals that our ability to separate sounds from background noise is affected by many variables, such as the gender of the speaker, the direction from which the sound is coming, the pitch and the rate of speech etc. Broadbent & Ladefoged (1957) adopted a systematic methodology to investigate the cocktail party problem. In their experiment they focused on, formant frequencies of vowels such as /i/ and /a/ perceived simultaneously by a listener. Liberman (1957) and his colleagues at Haskins laboratories conducted an experiment on human perception. They constructed a speech synthesizer to produce various sounds continuously from /ba/ to //da/ to /ga/. A group of listeners were
asked to identify as to which sound they heard and to discriminate between two different sounds. The results of the experiment showed that irrespective of the continuous variation of the sounds, listeners grouped sounds into discrete categories.

The Motor Theory of Speech Perception, is the hypothesis (Liberman, 1967), that people perceive spoken words by identifying the vocal tract gestures with which they are pronounced rather than by identifying the sound patterns that speech generates. Later on, Liberman & Mattingly (1985) revised this postulate to the intended gestures of the speaker (rather than the listener) in light of large data gathered by the researchers at Haskins Labs.

In 1952, Davis, Biddulph, and Balashek of Bell Laboratories built a system for isolated digit recognition for a single speaker, using the formant frequencies measured during vowel regions of each digit. They demonstrated as to how separate synthetic formants fuse to sound like a single vowel only when they have the same fundamental frequency (Fo). Miura and Koshikawa (1955), made a spectral study on Japanese vowel. A similar type of spectral study on the American English vowels was made by Potter and Steinberg (1950), to identify the positions of formants (first and second). Olson and Belar (1956) designed a system to recognize selected syllables for a single speaker at RCA Laboratories (Radio Corporation of America). In 1959, at University College in England, Fry and Denes
(1959) used a spectrum analyzer and a pattern matcher to recognize four vowels and nine consonants.

A major achievement of 1960's was the use of dynamic tracking (Vintsyuk, 1968) of phonemes in continuous speech recognition. Reddy (1966), Itakura (1975), extended the recognizer by identity of the speech and the speakers, including sex, by applying the linear predictive coding (LPC) technique. Later on, to design a perfect recognition system, clustering techniques were used to determine the number of distinct patterns required to represent all variations of different words across a wide user population.


These theories had a positive influence on the development of the phonetic research, initiating awareness of non phonological influences on speech perception, from word frequency to morphology and meaning, and finally to the successful machine speech recognition systems based on HMMs (Hidden Markov Model) and neural networks. They encouraged phonetic theories to replace binary
distinctions with probabilities, and speech perception as a probabilistic pattern-matching task. HMM is one of the key technologies developed in the 1980s (Ferguson, 1980; Rabiner, 1989, 1993). It is a doubly stochastic process which as an underlying stochastic process that is not observable (hence the term hidden), but can be observed through another stochastic process that produces a sequence of observations. Neural networks were first introduced in the 1950s, but had many practical problems. In the 1980s however, a deeper understanding of the strengths and limitations of the technology was achieved (Lippmann, 1987; weibel,1989; Katagiri et.al, 2003).

Recent cross-linguistic work (Beddor et al., 2001, 2002; Bradlow & Bent, 2002), examine the influences of native language structure on perception. Allen & Miller (2004) showed that listeners can identify the speaker from distinctions in VOT (Voice Onset Time). Most pertinently, Hawkins (2001, 2003); Smith (2004) showed that fine phonetic detail can facilitate identification of words in connected speech. Even inappropriate allophones in a sentence disrupted words spotting only when the speaker was familiar to the listener, even though all speakers used the same regional accent.

India has about 1652 dialects/native languages. Study on the acoustic features of some Indian language like-Tamil (Thananjayarasingham, 1988; Balasubramaniam,1972), Telugu (Nagamma,1981,1987), Malayalam (Velayudhan,
In view of the recent trend in the development of the speech technology preferably in the local languages, till now not much work has yet been done on any native language of the North-East region, based on proper scientific platform. In case of Assamese language, Basanti Devi (1984) carried out a study on the impact of voice on the length of vowels with the help of spectrographic analysis. Talukdar P.H. et al. (2004, 2005, 2006) had made a scientific study on Assamese, Bodo and Rabha languages.

1.3 OBJECTIVES OF THE STUDY

The strongest appeal of the 21st century is the “Communication without language barrier”. The most preferred means of communication amongst the people is the regional languages. The constitution of India recommended 22 languages as official languages, which are nearly in ten different Indian scripts. Most people enjoy speaking, reading or listening in one’s native language instead of English—the de facto language of interaction ranging from domestic matters to international affairs. But, to address any local problems, including agriculture, forestry, fishery, professionals or industry, it is imperative to give more focus on the use and development of local languages.
The North-Eastern region of India, called ‘Miniature India’, comprising of eight states, is an important geographical entity. The state of Assam has more than hundred tribes and sub-tribes, speaking some 30 to 35 dialects. Assamese, Hindi, Bodo and Rabha are the lingua-franca of the state. Bodo and Rabha are the dominant tribes of Assam. Linguistically, Bodo and Rabha belong to the Tibeto-Burman family of languages. The Bodo people from the Assam valley are mostly bilingual. The non-Bodos, living in the Bodo dominated territory prefer Assamese, Hindi or Bengali as their communicating medium. It seems that due to some intrinsic characteristics of the Bodo language, the language is facing lack of popularity among the other communities of the state. This has created a barrier, not only in the process of mutual understanding between Bodo community and other linguistic groups, but also, in all other related domain such as social, in developing special relationship, cultural exchange, education etc. This has resulted some kind of stagnation of Bodo vocabulary in terms of its popularity. This poses as a serious drawback in the all round development of the Bodo language. This might be due to lack of well-structured formulation of the language as far as national and international standards are concerned. However, in a cross regional context, a Bodo person fully cannot enjoy the sense of belonging, faces obstructions in exchanging ideas, also in developing friendship with other communities.

According to our study, the Rabhas, who were once considered as the ancestors of many sub tribes of North-East region have eleven linguistic sub-
groups, out of which only three of which have their own writing system. However, all the sub-groups have been now unified so as to identify into commonly accepted form of the language i.e. Rangdani, using Assamese script. This would facilitate the process of integration with Assamese society. At present, their common media of conversation is a broken variety of Assamese and Bodo languages mixed with Rabha phonemes.

In view of the above circumstances, the present study is mostly concentrated on proper analysis and representation of the Bodo and Rabha languages. Considering all the linguistic and non-linguistic features of the languages it is expected that the study would reveal some hidden characteristic of both the languages which would strongly advocate its identity and uniqueness with respect to other tribal languages of the North-Eastern region of India.

Some groups of speech researchers in India have already made remarkable contribution towards the development of speech technology in local languages like Bengali, Malayalam, Tamil, Telugu etc. It is the aim of the present study to develop a role-model for Bodo and Rabha language which would be able to address some common linguistic problems of these two languages and enable both the communities to enjoy a sense of integrity & homogeneity in the entire scenario of the North-East region.
1.4 ORGANIZATION OF THE THESIS

The results of the present research work are presented in nine chapters as mentioned below:

CHAPTER-I: This chapter gives an introduction to the speech research. Apart from outlining the objective of the present study, a survey of past works on human speech is also been included.

CHAPTER-II: A brief illustration of the history of Bodo and Rabha languages with their phonological representation has been included in this chapter.

CHAPTER-III: This chapter is devoted to a simple introduction to the human speech, anatomy and physiology of the human speech production system and speech modeling techniques.

CHAPTER-IV: An acoustical analysis of speech signal has been detailed in this chapter. The analysis is based on the Linear Predictive Cepstral Coefficients (LPCC) and Formant frequency estimations and their relevance in the present study is thoroughly discussed. The characteristic features extracted from the Bodo and Rabha vowels and some phonetically balanced open and closed syllabic word sets are also taken for the investigation, including for both the male and female speaker and their range of variations are investigated.

CHAPTER-V: In this chapter the intonation pattern of the Bodo and Rabha languages have been studied in details. Here, the variation of tone of Bodo and
Rabha word sets are investigated with the help of corresponding pitch contour. The tonal representation have been marked with specific symbols, such as: for rising (>), falling (<) and flat (−).

CHAPTER-VI: This chapter is devoted to the introduction of prosody variation of the Bodo and Rabha languages. A careful observation is made to observe as to how the meaning of the same sentence changes with the change of prosody of the sentence.

CHAPTER-VII: This chapter is devoted to the word Boundary detection of Bodo and Rabha sentences. Phonetically balanced selected sentences are analyzed in time and frequency domain and the pitch contour of the respective sentences are thoroughly discussed. The pitch variation across a word in a sentence is measured and the differences of the last pitch value of a word and the first pitch value of next word in a sentence is applied for word boundary detection.

CHAPTER-VIII: In this chapter the recognition of the segmented word has been done for selected words. The LPC coefficients extracted from the selected word sets are subjected to the vector quantization (VQ) and K-means algorithm for pattern matching.

CHAPTER-IX: This chapter contains the overall conclusion of the present study and discusses the scope for future work.