

NEED, PROBLEM, PROSPECT AND POTENTIAL OF SPOKEN LANGUAGE TECHNOLOGY IN INFORMATION RETREIVAL FOR DEVELOPING COUNTRIES

Asoke Kumar Datta
Computer Vision and Pattern Recognition Unit
Indian Statistical Institute,
203 B T Road
Kolkata 700035
India

1.0 INTRODUCTION

Knowledge is power. It is imperative to harness this power in the form that may be used for national development. Thanks to the tremendous development in the information technology in the last few decades, the accrual of knowledge at the global level in respect of all conceivable disciplines relevant to human development, starting from archeology (beginning of civilisation) to cosmology (end of creation) may be considered to be adequate. It is at least so for the development of standards of life in the developing countries in south Asia. All this knowledge is now available at the press of few buttons on the computer. There may be a need for the development of the e- infrastructure to particularly cater for proper utilisation of the existing knowledge to help humanity in developing countries:

- to deliver real time information and customise knowledge to improve decision making for end-users in market demand, quality, productivity, price, technology, etc.
- to aggregate demand in the nature of producer's e-Co-operatives to access Quality products at low cost
- to set up e-Commerce link to reduce wasteful intermediaries, multiple handling, transition cost etc. and to make logistic efficient, cost effective.

However paying attention to these alone may prove to be futile as the end users for national development will still remain deprived of the benefit of it. It is so and shall remain so as all the knowledge and technology is in an alien language. In my view for real development of the human beings in these countries the end users are the grass root workers like those in agriculture, aqua-farming, poultry, animal-husbandry, forest-produce, cottage industries like ceramic, handicraft, leather, rubber, weaving etc. While localisation may be helpful, yet, I am afraid, this help would only be marginal. As will be explicated later the main reasons are two. One is illiteracy. The other one is that even for those who are functionally literate it may be an uphill task, except for the few highly educated people (most them may not even have a need for it), to get them motivated to acquire sufficient e-skill for operating search devices.

One must note that the most common and handy medium for knowledge transfer amongst people is the speech mode. In the present context, for developing countries, this has an extra significance compared to the developed countries. In developed countries, for better or worse, speech is becoming marginal any way. In fact, e-chat seems to be more functional than voice-chat there. Knowledge is more easily acquired impersonally through reading media. However, for developing countries this oral transmission of knowledge is still the force majore.

Traditionally, in India, knowledge is transferred and readily consumed through personal direct interaction between teacher and student. The primary objective of this presentation is to attract attention to Spoken Language Technology (SLT) as the most favoured means for knowledge

transfer to achieve development of mankind in the developing and underdeveloped countries. As we shall see later that India may be considered to be at a stage ready for an effective take-off towards achieving technology for oral transmission of the e-knowledge to the common mass. Apart from mass education programs, its impact on national production and commerce appears to be tremendous.

2.0 NEED FOR SLT

Let us look at the language profiles, population and literacy rates prevalent in some of the south Asian countries with particular reference to India. Table 1 gives currently available figures for south Asian countries. In most cases the literacy rates are compiled through general census. These are therefore casual in the sense that these are based on mere response of the respondent. There is no scope for rigorous verification. The literacy rate generally reflects the ability to read and write. The functional literacy therefore is likely to be quite lower than the given figures. The main points to note are two. One is that the knowledge as it now exists is in western languages, particularly English and that 92% of the population of this region would therefore be denied access to this knowledge. The other point is that even after a complete localisation is achieved the written knowledge will escape access by almost over 60% of the population in the region just because they will be unable to comprehend the written text. The point is that while localisation is

Table 1. Population and literacy profile for south Asian countries

COUNTRY	POPULATION	LANGUAGE	LITERACY	ENGLISH LITERACY
Bangladesh	123,062,800 (1996)	Bangla (official)	NA	23%
Bhutan	1,822,625 (1996)	Dzongker (Official) Tibetan Dialects Nepalese Dialects	24%	NA
India	1,033,000,000 (2001)	Sanskrit Hindi Urdu English	65.38%	8%
Maldives	270,758 (1996)	Muldivial Divehi	53%	NA
Nepal	22,367,048 (1999)	Nepali	NA	NA
Pakistan	124,540,000 (1999)	Urdu (National) English (Official)	38.9%	NA
Srilanka	18,129.830 (1994)	Sinhala (official) Tamil (National)	65%	10%

absolutely necessary as otherwise even verbalisation would be meaningless for the mass, one just cannot stop there. Also in my belief human development need not await till a 100% literacy rate is achieved.

Table 2 gives the literacy rates in the eastern region of India which is slightly better than the overall figures for south Asian region. Except for Bihar and Jharkhand and including Bangladesh the people in these region speak languages which have close similarity with Bangla. This calls for an immediate attention to the development of spoken language technology in this language. One may also note here the prevailing human ethos in this region wherein emphasis is given on

oral transmission of knowledge as opposed to the western ethos of receiving knowledge from text. There exists a positivity towards understanding through hearing rather than through reading amongst the common mass. This may be due to massive illiteracy but none-the-less this is a fact. The extensive TV coverage existing in India provides a direct opportunity to massive flushing of knowledge in all conceivable aspects of human development through the use of non-interrogative forms of SLT which can be developed fast.

Table 2 Literacy rates in the Eastern Region of India

West Bengal	69.22%
Assam	64.28%
Orissa	63.61%
Bihar	47.53%
Jharkhand	64.13%
Tripura	73.66%
Manipur	68.87%

3.0 PROBLEMS

Let us now look at the non-technological problems related to knowledge communication in speech mode. Except India we have to tackle eight different languages even if we, for the present, shut our eyes on dialectical variations. Table 3 presents the minimum need for India. We need to address 17 different languages. It is said there are around 4000 dialects including tribal languages (for most of which there is no script).

Table 3. Official regional languages in India

Official National		Sanskrit Hindi Urdu English
Regional	North	Kashmiri Panjabi
	West	Gujrati Marathi Konkani
	East	Bangla Asamiya Oriya Nepali Manipuri
	South	Tamil Malayali Kanada Telugu

The problem is compounded because of the absence of comprehensive research materials in most of the languages. Altogether there is a need to tackle about 22 different major spoken languages. I believe they could be grouped under three major headings. Though the acoustic phonetic structures under each group may exhibit some sort of similarity the acoustic prosodic structures are likely to be quite distinct. In India during the last two decades some studies have been reported on acoustic phonetics of some Indian spoken languages, e.g. Hindi, Bangla, Telugu, Tamil, Kanada, Malayali, Assamese, And Oriya. However most of them are local efforts and therefore lack standard formats. Speech corpora is almost non-available in most of the languages. In fact, I am not aware of a central focus on development of corpora in major Indian languages needed for IT development. Apparently the central language planners are unaware of the need and the dimensions of the total problem. While there may be funds available with the central agencies the spoken languages failed to attract the needed attention. During the mid-80s to early-90s there has been some attention towards speech technology through the KBCS (Knowledge Based Computing Systems/ Fifth generation Computers) programs. These, inter alia, produced some

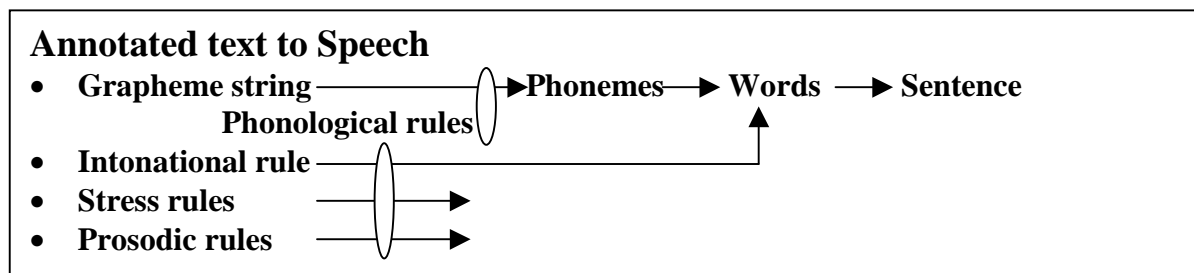
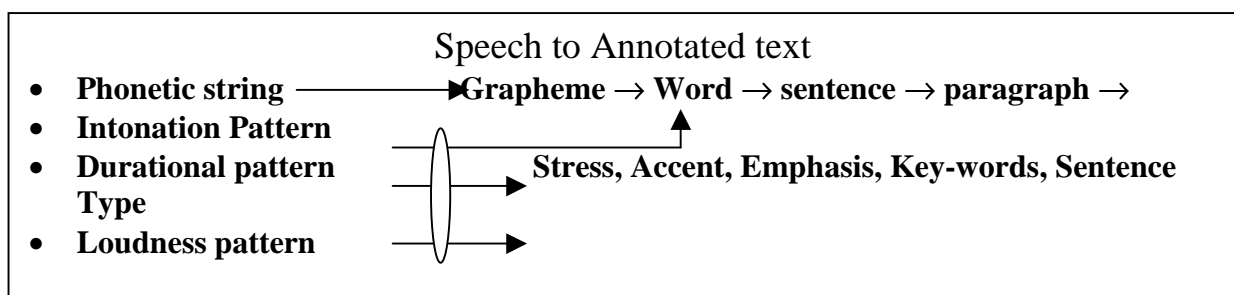
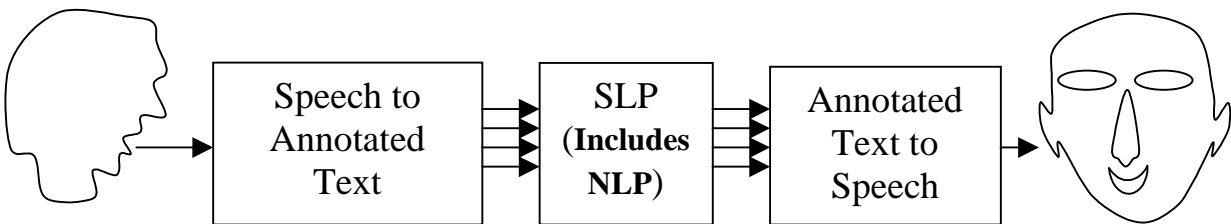
high level human resources at a few institutions. During the last 10 years the focus was allowed a natural death.

The present need seems to be twofold. One is the establishment of at least four regional centers for SLT throughout India with adequate number of linguistics and software people. The other is the development of human resources through the conduct of crash courses for these centers on speech technology, and signal processing. The objective of regional centers would be development of speech corpora for the relevant regional languages as well as development of acoustic phonetic and acoustic prosodic databases. The technology for knowledge retrieval can be developed in some of the existing centers of speech research in India.

4.0 PROSPECT AND POTENTIAL

Speech research in relation to artificial intelligence dates back to late sixties in India. In fact the first fortnight long workshop was held under the guidance of Prof. Rais Ahmed in Aligarh Muslim University sometimes in 1968. In a sense this may well be considered as an effort towards a consolidation of the beginning of speech technology research in India. This research got a big boost during the period from mid-80s to mid-90s and several centers in different parts of the country became actively engaged in various aspects of SLT. Studies on acoustic phonetics in Hindi, Bangla, Assamese, Oriya and Telugu were actively pursued and some statistical data bases were developed. Research on speech recognition included studies on efficient parameters, both in time domain and signal domain, and classificatory analysis using various statistical

Figure 1. Basic elements of an SLP system



decision theoretic, fuzzy set theoretic and neural net approaches. These studies were made on at least three major Indian spoken languages, namely, Bangla, Telugu and Hindi. Indigenously developed limited vocabulary word recognition were demonstrated in late eighties. Bangla text-to-speech synthesizer appeared in the beginning of nineties. Some premier scientific and academic institutions took up specific courses of study, research and training in signal processing as well as physical, cognitive and engineering aspects of speech research. In short, in my opinion, India is poised for a take-off into meaningful SLT development.

Let us briefly examine the task of SLT in the present context. This may be illustrated as in figure 1. At the front end is the device, which converts speech signal into annotated text. This essentially selects and converts appropriate segments of the electrical signal from the microphone and finally converts, using certain extracted parameter and decision theoretic tools, first into phoneme strings (alternatively directly into words) then into graphemes, words, clauses and sentences. In the process it may have to use lexical knowledge. Annotation refers to providing stress, emphasis, target words and the like. Though simply put the actual process is quite complex.

The basic purpose of the Spoken Language Processing (SLP), in the present context is to analyse the annotated text to understand the nature and target of the queries on one hand, and on the other hand to locate, extract and organise the required answers later. The annotated text is essentially a graphemic representation of the spoken passages along with the stress, accent, emphasis, etc. and the declination resets for aiding detection of clausal and phrasal boundaries. This aided with an NLP is likely to extract, at a reasonable level, the nature and targets for the queries. A good NLP would be further required to organise the answer from the stored knowledge bases.

At the rear end is a TTS which converts the written text into natural speech. Before we begin to elaborate the complexities and the studies and researches involved in fulfilling the objective, even in a limited task domain an examination of the present status of technology available in the country is in order.

4.1 Speech To Text

A list of organisations and institution wherein speech research and technology are being pursued seriously in a broad spectrum would be presented in a later section. Here we shall briefly enumerate the results of the latest study conducted in Indian Statistical Institute (ISI) in mid-90s. The study was made in the context of a general Reservation Enquiry System using Speech Mode Interaction (RESMI). It was basically a Manner-Based Lexically-Driven Word Recognition System. The spoken word is converted into a corresponding pseudo-word into four manner-based symbols, namely fricatives, stops, partially obstructed vocalics and vowels. A study with 20,00 strong Bengali lexicon revealed that these pseudo words divided the lexicon with an average cohort size of 4.5. There was about 8000 single cohort for which no further disambiguation is necessary for word recognition. A current study being conducted at ISI indicates that a phase space base analysis can provide the manner based labeling with 97.5% accuracy. Earlier experiments conducted in ISI attained recognition rate of 83% for steady state vowels in word context and 75% recognition rate for plosives. A lexically driven acoustic expert system was also tested on the same vocabulary after partitioning using the manner based labeling. The system looks at the different vowel positions and tries to find which position can be used to further make a binary sub-partition in the cohort such that the two sub-partitions may contain non-contiguous vowel groups like /a/, /u/, /o/ in one group and /e/, /i/ in another group It

may be noted that vowel classification results approach nearly 99% in non-contiguous groups. This process is continued till all words in the group are uniquely partitioned. This whole knowledge is consolidated in the lexically driven acoustic expert system. The average attainable recognition rate was estimated for the whole vocabulary using the detailed knowledge of vowel and consonant classification obtained from earlier experiments with spoken data base in three different Indian languages. Interestingly for only 7% of the words consonant classification was necessary for complete disambiguation. An estimate showed a word recognition rate of 95%. Investigation also revealed a number of shape domain parameters which were as efficient as the spectral parameters but were simpler and more robust for extraction. However use of this approach requires accurate and robust word-segmentation procedure. Word boundary detection from continuous speech is yet a difficult task.

4.2 Text-To-Speech Systems

At present TTSs in at least three Indian languages are reported to be developed in the country. ISI developed indigenously in 1990-91 BANGABANI, which produces Bangla, Hindi and Indian-English continuous flat speech using ESOLA technique. There also exists parametric synthesiser developed later for Hindi and Hinglish. Very recently the development of a diphone based Malayali synthesiser is also reported. ISI is currently engaged in developing a modified version of the earlier concatenative synthesis engine to produce properly inflected natural like 8 KHz bandwidth speech with signal a dictionary approximately of 50 Kbyte. It would then be possible to include TTS in a local language in a simputer. However one must make a note at this point that the development of synthesis engine is only the tip of the total iceberg of natural speech synthesis. An extensive and long time research programs are needed to be undertaken for the development of necessary linguistic infrastructure and knowledge bases in each of the local languages.

5.0 RESOURCE FOR SLT

While a significant amount of activities is noticed in India in the field of research and corpora development in linguistics, this is primary limited to text and NLP. The national focus so far is limited to the development of resources in regional languages in this domain. For a brief period of about a decade from mid-80s to mid-90s some sporadic but intense activities in limited pockets took place in the area of development of SLT under the general ambit of development of Knowledge Based Computing Systems. Because of this and because of the activities for about four decades in ISI in the area one may draw some idea of the dimension and magnitude of an effective policy for the development of SLT in India. At the national level the programme requires a simultaneous effort to tackle 17 languages spread over the whole country. Which means regionalisation of the programs under a central control. Regional centers need to be developed with adequate IT infrastructure, linguists and speech scientists. Apart from conducting basic research on acoustic phonetics and acoustic prosodics the major task is the development of spoken language corpora for formation of knowledge bases for e-engines required for SLT. Table 4 gives a very preliminary list of requirements for the purpose. Analysis here refers to that in terms of speech recognition. Each item in the third column may require a time span of 2 – 3 years to complete the study and analysis. The items under synthesis is listed keeping in mind the signal domain concatenative approach as it provides quicker solution and natural-like synthetic

Table 4. A general view of minimal corpora for SLT in one spoken language

Analysis	Phonetics	Phonetically balanced words embedded in neutral sentences	≈500	Digital Audio Recording male/female/child 5 each
		Sentences for word juncture	≈100	
	Prosodics	Text reading	3x10 minutes	DAR 5 male 5 female
		Querries	50 sentences	
		Dialogues	6x5 minutes	
	Synthesis	Phonetics	Tetra-syllabic non-sense words	≈500
Prosodics		Sentences to cover all general intonation patterns	≈ 100	

speech. The development of the signal dictionary for concatenative synthesis may be completed within one year provided basic phonetic structure of the language is known. For languages with complex phonology the task of text-to-phoneme conversion may take some time if electronic dictionary is not available. Analysis of spoken sentences for the development of intonation and prosodic rules again may be time consuming.

Some speech research centers in India

Indian Statistical Institute, Sangeet Research Academy,

Jadavpur University, E R D C I (all Calcutta)

CEERI (Delhi)

Aligarh Muslim University (Aligarh)

CIIL, AIISH (Mysore)

TIFR (Bombay)

IIT (Madras, Guwahati)

Osmania University (Hyderabad)

ISDL (Thiruvananthapuram)