



SCIL - Speech Corrector for Indian Languages -An Algorithm to Improve Speech Recognition Accuracy

Dr.K.V.N.Sunitha

Principal, BVRIT Hyderabad,
College of Engg. for women Bachupally,
Hyderabad,A.P., India

A.Sharada

Associate Professor, CSE Dept,
G.Narayanamma Inst of Tech & Science
Hyderabad, A.P., India

Abstract - Even after years of extensive research and development, accuracy in ASR (Automatic speech recognition) remains a challenge to researchers. The reason is the inherent complexity of inflectional languages, and lack of resources. When processing languages with extremely rich word forming, the resulting word lists are typically very large, which is demanding from a computational point of view. A more serious problem is that many perfectly valid word forms are likely to be missing from the list anyway, since they might never have occurred in the corpus used as a source. For words which are out of the dictionary the recognition accuracy is quite low and gets matched to the nearest possible word in the dictionary. In this paper we propose method that enhances the speech recognition accuracy for inflectional languages in general and, Indian languages in particular.

Keywords: Syllable Identifier, Phone-SequenceGenerator, Morph analyzer, Segmenter, Inflection grabber

I. INTRODUCTION

Speech technologies are of particular importance to individuals with physical impairments that hinder their use of traditional input devices such as the keyboard and mouse. Even after years of extensive research and development, accuracy in ASR (Automatic speech recognition) remains a challenge to researchers. There are number of well known factors which determine accuracy. The prominent factors include variations in context, speakers and noise in the environment. Therefore research in automatic speech recognition has many open issues with respect to small or large vocabulary, isolated or continuous speech, speaker dependent or independent and environmental robustness. The accuracy and acceptance of speech recognition has come a long way in the last few years and forward-thinking contact centre operations are now adopting this technology to enhance their operation and improve their bottom-line profitability. The performance of speech recognition systems is usually specified in terms of accuracy and speed. Accuracy may be measured in terms of performance accuracy which is usually rated with word error rate (WER), whereas speed is measured with the real time factor. Other measures of accuracy include single word error rate (SWER) and Command success rate (CSR).

Most speech recognition users would tend to agree that dictation machines can achieve very high performance in controlled conditions. For simple applications training of the acoustic models usually require only a short period of training and may successfully capture continuous speech with a large vocabulary at normal pace with a very high accuracy. An accuracy of 98% to 99% can be achieved if operated under optimal conditions. 'Optimal conditions' usually assume that users:

- have speech characteristics which match the training data,
- can achieve proper speaker adaptation, and
- Work in a clean noise environment (e.g. quiet room).

This explains why some users, especially those whose speech is heavily accented, might achieve recognition rates much lower than expected. Limited vocabulary systems, require no training, can recognize a small number of words (for instance, the ten digits) as spoken by most speakers. Such systems are popular for routing incoming phone calls to their destinations in large organizations.

A. Word Structure of Inflectional Language

Inflectional language is characterized by a rich system of inflectional morphology and a productive system of derivation, saMdhI (conation of full words) and compounding. This means that the number of surface words will be very large and so will be the raw feature space, leading to data scarcity[1].

In inflectional language every word consists of one or several morphemes into which the word can be segmented; consider for instance the morpheme segmentations of the following Telugu words: "Ame(she), Ame+yokka(of her), Ame+tO(with her), AmE+nA(is it she)". In highly-inflecting and compounding languages the number of possible word forms is very high. This poses special challenges to NLP systems dealing with these languages. For example, in automatic speech recognition it is customary to use pre-made lists of attested word forms as a "normative" vocabulary.

The incoming acoustic signal is matched against the list, and only words contained in the corpus can be recognized. Such a word list can be created by collecting word forms from large text corpora or existing lexicons, and the aim is to obtain as much coverage as possible of the words of the language. When processing languages with extremely rich word forming like Telugu, the resulting word lists are typically very large, this is demanding, from a computational point of view. A more serious problem is that many perfectly valid word forms are likely to be missing from the list anyway, since they might never have occurred in the corpus used as a source. For words which are out of the dictionary the recognition accuracy is quite low and gets matched to the nearest possible word in the dictionary.

II. LITERATURE SURVEY

Most of the existing ASRs correct the errors based on context. They use phone, syllable or word level modeling and based on n-gram statistics they correct the word. Phone level is difficult to implement because of co-occurrence effects of phones. Telugu is morphologically very rich language. Word level cannot be used, as it is impossible to cover all the words of the language because of its large vocabulary and changes in the word forms. So we use morphemes as basis for error identification. Factored language models have recently been proposed for incorporating morphological knowledge in the modeling of inflecting language. As suffix and compound words are the cause of the growth of the vocabulary in many languages, a logical idea is to split the words into shorter units [2].

III. PROPOSED MODEL

In inflectional language every word consists of one or several morphemes into which the word can be segmented. The approach used here aims at reducing the above mentioned problem of having a very huge corpus for good recognition accuracy. It exploits the characteristic of Telugu language that every word consists of one or several morphemes into which the word can be segmented.

A. Architecture of the Proposed Model

The design of Speech Corrector for Indian Languages, consists of the Syllable Identifier, Phone Sequence Generator, Word Segmenter, and Morpho Syntactic Analyzer modules. Input speech is decoded by a normal ASR system which gives the identified word as a string. The sequence of phones would be the input to the **Word Segmenter** module which matches the phonetized input with the root words stored in dictionary module, and generates a possible set of root words. **Morpho-Syntactic Analyzer** compares the inflection part of the signal with the possible inflections list from the database and gives correct inflection. This will be given to Morph Analyzer to apply morpho-syntactic rules of the language and gives the correct inflected word.

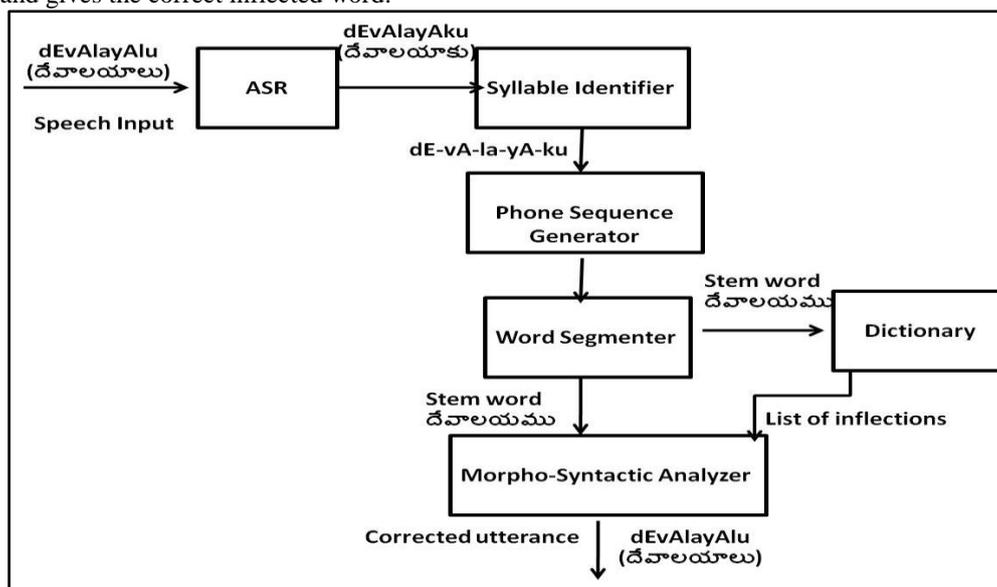


Fig 1: Block diagram of proposed model

i) Syllable Identifier

Syllable identifier marks the rough boundaries of the syllables and labels them. At this stage, we get list of syllables separated with hyphen. The user input is syllabified and this would be the input to the next module. E.g. dE-vA-la-yA-ku

ii) *Phone Sequence Generator* As the words in the dictionary are stored at phone level transcription, this module generates the *phone sequences from the syllables*. E.g. d-E-v-A-l-a-y-A-k-u

iii) *Word Segmentor* This module compares the phonetized input from starting with the root words stored in dictionary module and lists the possible set of root words. The possible root word is *dEvAlayamu*.

iv) *Dictionary* Dictionary contains stems and inflections separately. It does not store inflected words as it is very difficult, if not impossible, to cover all inflected words of the language. The database consists of 2 dictionaries:

- a) Stem Dictionary
- b) Inflection Dictionary

Stem dictionary contains the stem words of the language, signal information for that stem which includes the duration and location of that utterance and list of indices of inflection dictionary which are possible with that stem word.

Inflection Dictionary contains the inflections of the language, signal information for that inflection which includes the duration and location of that utterance.

Both the dictionaries are implemented using trie structure in order to reduce the search space. Details of this implementation can be seen in [3].

TABLE 1: STEM DICTIONARY

| Word | Stem signal information | Inflection Indices |
|-------------------|-------------------------|--------------------|
| amma (అమ్మ) | D:\work\s1.wav | 1, 2, 4, 6,7,8 |
| anubhUti(అనుభూతి) | D:\work\s2.wav | 1,7 |
| ataDu (అతడు) | | |

TABLE 2: INFLECTION DICTIONARY

| Sl.No | Inflection | Inflection Signal Information |
|-------|-------------------|-------------------------------|
| 1 | ki (కీ) | D:\work\inf1.wav |
| 2 | tO (తో) | D:\work\inf2.wav |
| 3 | guriMci (గురించీ) | |
| 4 | | |

v) *Morpho Syntactic Analyzer* This module compares the inflection part of the signal with the possible inflections list from the database and gives correct inflection. This will be given to Morph Analyzer to apply morpho-syntactic rules of the language and gives the correct inflected word.

IV. SPEECH CORRECTION PROCEDURE

Speech correction procedure is explained below:

1. Capture the utterance.
2. Get its syllabified form.
3. Generate phone sequence from the syllabified word.
4. Compare the phone sequences with stem words in the dictionary and identify the stem.
5. Segment the word into stem and inflection.
6. Get the list of possible inflections.
7. Compare the inflection signals possible with that stem one by one and apply morpho-syntactic rules of the language to combine stem and inflection.
8. Display the inflected word.

Using the rules the possible set of root words are combined with possible set of inflections and the obtained results are compared with the given user input and the nearest possible root word and inflection are displayed if the given input is *correct*. If the given input is *not correct* then the inflection part of the given input word is compared with the inflections of that particular root word and identifies the nearest possible inflection and combines the root word with those identified inflections, applies sandhi rules and displays the output. When there is more than one root word or more than one inflection has minimum edit distance then the model will display all the possible options.

User can choose the correct one from that. E.g., when the given word is *pustakaMdo* (పుస్తకండ్), the inflections *tO* making it *pustakaMtO* (పుస్తకంతో) meaning 'with the book' and *IO* making it *pustakaMIO* (పుస్తకంల్) meaning 'in the book') mis are possible. Present work will list both the words and user is given the option. We are working on improving this by selecting the appropriate word based on the context.

Algorithm SCIL:

1. W=Utterance.wav
2. Syl[]=SyllableIdentifier(W)
3. Phone[]=phonetizer(Syl[])
4. Stem=getStem(Syl)
5. Infl[]=getInflections(Stem)

6. While (not exactMatch)
word=MorphAnalyzer(stem,inflMatch)
7. display word
8. Stop

V. EXPERIMENTAL RESULTS

The approach is tested using 1500 speech samples. These samples consist of 100 distinct words, each word repeated 3 times and recorded by 5 speakers in the age group 18-50. It is implemented as a speaker dependent system. An average model is built from the three utterances of each word for each speaker. Each speaker is given a unique ID, using which average model of that speaker is used for testing.

TABLE 3: SCIL RESULTS

| | |
|---|-----|
| No.of Distinct words | 100 |
| No.of Speakers | 5 |
| No.of Test samples | 500 |
| No.of words correctly recognized | 299 |
| No.of words misrecognized at Inflection | 157 |
| No.of words misrecognized at Root | 74 |
| No.of words corrected by SCIL | 130 |

VI. CONCLUSION

Main focus of the work presented here is to experiment whether morphology based recognition help in speech correction. So, it is implemented as a speaker independent and gender specific system. From the work we observed that speech recognition errors are more at sandhi formation meaning that inflectional languages will have more Word Error Rate. Through the SCIL algorithm we are able to correct up to 90% of the errors occurring at inflections.

REFERENCES

- [1]. G. Clopper, Cynthia G., Pierrehumbert, Janet B. / Tamati, Terrin N. "Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise", *Laboratory Phonology*. Volume 1, Issue 1, Pages 65–92, ISSN (Online) 1868-6354, ISSN (Print) 1868-6346
- [2]. Lamel and G. Adda, "On Designing Pronunciation Lexicons for Large Vocabulary, Continuous Speech Recognition", Proc. International Conference on Spoken Language Processing (ICSLP'96), pp6-9, 1996.
- [3]. Dr.K.V.N.Sunitha, A.Sharada," A Novel approach to overcome data scarcity problem for highly inflectional languages", International Conference on Recent Trends in Information, Telecommunication and Computing – ITC 2010 held in March 2010, pp 290-292 Available in ACM Digital Library
- [4]. M. Sigman 1, 2 and G.. Cecchi, "Global organization of the lexicon", Proceedings of the National Academy of Sciences, Feb 2002, Vol 99, No.3
- [5]. G.Uma Maheshwara Rao, "Morphological complexity of Telugu", ICOSA A -2, 2000.
- [6]. J.Lovins. "Development of stemming algorithm", Journal of mechanical translation and computational linguistics, 11:22-31,1968
- [7]. K.V.N.Sunitha, A.Sharada, "Building an Efficient Language Model based on Morphology for Telugu ASR", KSE-1, March 2010, CIIL, Mysore
- [8]. K.V.N.Sunitha, A.Sharada, "Telugu Text Corpora Analysis for Creating Speech Database", IJEIT,ISSN 0975-5292, Dec 2009, Volume 1, No.2
- [9]. Paice C and Husk G. "Another Stemmer". In ACM SIGIR Forum 24(3):566,1990
- [10]. M.F.Porter. "An algorithm for suffix stripping".In readings in information retrieval, pages 313-316, San Francisco,CA,USA,1997.Morgan Kaufmann Publishers Inc.
- [11]. Jinxi Xu and W. Bruce Croft."Corpus based stemming using co-occurrence of word variants". ACM Trans.Inf.Syst., 16(1):61-81,1998
- [12]. vishwabharat@tdil
- [13]. J.L.Dawson. "Suffix removal for word conflation". In Bulletin of the Association for Literary and Linguistic Computing, volume 2(3), pages 33-46, Michaelmas,1974
- [14]. R.Krovetz. "Viewing morphology as an inference process". In Proceedings of Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 191-203, 1993
- [15]. Emerald Group Publishing Ltd., "Issues in Indian languages computing in particular reference to search and retrieval in Telugu Language"
- [16]. "LANGUAGE IN INDIA Strength for Today and Bright Hope for Tomorrow", Vol 6, August2006.
- [17]. K.Nagamma Reddy, "Phonetic, Phonological, morpho-syntactic and semantic functions of segmental duration in spoken Telugu:acoustic evidence"
- [18]. Dr.K.V.N.Sunitha, A.Sharada, "Thesaurus based web searching", Springer Publications, Advances in Intelligent and Soft Computing, 2012, Vol. 166/2012, 265-272, DOI:10.1007/978-3-642-30157-5_27, ISSN: 1867-5662

- [19]. Dr.K.V.N.Sunitha, A.Sharada, "Spelling Corrector for Indian Languages", Springer Publications, Advanced Computing - Communications in computer and information science, 2011, volume 133, Part 5, 390-399, DOI:10.1007/978-3-642-17881-8_37, ISSN:1865-0929
- [20]. Dr.K.V.N.Sunitha, A.Sharada, "Dynamic Construction of Telugu Speech Corpus for Voice Enabled Text Editor", IJHCI, Vol 3., Issue 4, 2012, pp:83-95

AUTHOR PROFILE



Dr.K.V.N.Sunitha Currently working as Principal, BVRIT Hyderabad college of Engineering for women, Nizampet, Hyderabad has done her B.Tech ECE from Nagarjuna University, M.Tech Computer Science from REC Warangal. She completed her Ph.D from JNTU, Hyderabad in 2006. She has 21 years of Teaching Experience, worked at various engineering colleges. She received "Academic Excellence Award" by the management of G.Narayanamma Institute of Technology & Science on 18th September 2005. She also received "Best computer Science engineering Teacher award for the year 2007" by Indian Society for Technical Education ISTE. She has been recognized & invited by AICTE as NBA expert evaluator. Her autobiography was included in "Marquis Who is Who in the World", 28th edition 2011, since August 2012. She has authored three text books, "Programming in UNIX and Compiler design"- BS Publications & "Formal Languages and Automata Theory" by Tata Mc Graw Hill & "Theory of Computation" by TMH in 2011. She is an academic advisory member & Board of Studies member for other Engineering Colleges. She has published more than 65 papers in International & National Journals and conferences. She is a reviewer for many national and International Journals. She is fellow of Institute of engineers, Sr member for IEEE & International association CSIT, and life member of many technical associations like CSI and ACM.



Mrs.A.Sharada presently working as Associate Professor in CSE Dept, G.Narayanamma Institute of Tech& Science, Hyderabad. She has completed her M.Tech from JNTU Hyderabad and currently pursuing Research in the area of Telugu Speech Recognition. She is Associate Member of Inst.of Engineers India and Life Member of CSI.