DEVELOPMENT OF A SPEECH RECOGNITION SYSTEM FOR TAMIL FOR RESTRICTED SMALL TASKS

A.Nayeemulla Khan and B.Yegnanarayana

Department of Computer Science and Engineering
Indian Institute of Technology, Madras, 600 036, India
E-mail: {nayeem@, yegna@}speech.iitm.ernet.in

ABSTRACT

In this paper we describe an online speech recognition system for Tamil for tasks involving a small vocabulary (10-100 words). Two tasks are considered for illustration, namely, TV control (17 words) and banking application (61 words). The recognition system is based on Hidden Markov Models (HMM) and is implemented using Entropic's HMM toolkit.

1. INTRODUCTION

The objective of this study is to address issues in developing a simple dialog system with a machine for a restricted task in speech mode using a local language like Tamil. The two main components of such a system are speech recognition and a text-to-speech synthesis system. We will address the speech recognition component in this paper.

For any communication with a machine in speech mode, it is essential to define the task in the form of a grammar and vocabulary. For a restricted task the vocabulary can be limited to a small set of 20-100 words. The syntax for the grammar may be simple as there may not be many constraints on the sequence of words in an utterance for a restricted task. The most important issue in a dialog system is the ill-formed nature of the input utterance, which may not conform to the constraints of the grammar. We will not consider the issue of ill-formedness in this paper.

In order to build a speech recognition system it is necessary to define the basic subword units which provide a link between speech and text. Normally the basic units could be one of the following: phonemes, syllables or words. The smaller the unit, the more will be the difficulty in recognizing it in continuous speech, due to variability in the characteristics of the unit in the context. The larger the unit, the more difficult it will be to derive a suitable model to describe the speech utterance due to limited amount of data.

A standard approach of designing a speech recognition system is by using a Hidden Markov Model (HMM) for each subword unit [1] . Speech data for each subword unit is obtained using training data which is generated by reading out a list of training sentences derived from the grammar for a given task [2] . Each word in the vocabulary is expressed as a sequence of subword units, thus creating a pronunciation dictionary, for the words. The pronunciation dictionary together with the grammar, captures the pronunciation in terms of the subword units, for all possible sentences in the given task. The recognition of an unknown test utterance is based on determining the sentence that is most likely to have produced the observation sequence of symbols (subword units) corresponding to the test utterance. In the next section we will describe the task and the recognition system we developed for Tamil.

2. SPEECH RECOGNITION SYSTEM FOR TAMIL

We describe the development of speech recognition systems for two application areas. The first application is a command sequence likely to be used for voice-controlled Television (TV) that may be used by a paraplegic. This application uses a vocabulary of 17 words. Each utterance is about 2-4 words long. The second application is the likely dialog between a person and a machine in a banking environment. Here the speech contains a string of connected digits for the bank account, and an identification code, followed by instructions on the services required, which may also optionally contain digit sequences along with the start and ending delimiters for the transaction. Each utterance is about 20-30 words long.

Although the grammar for the input utterances in Tamil for these task is not well defined, we restrict the allowable sequence of words. The disadvantage of this approach is that even unrelated speech or nonspeech will be forced into one of the possible sequences of legal utterances. The goal is to estimate the probability of a word sequence for the recognition task. The grammar for the two tasks are listed in Tables 1 and 2. Throughout we have used ITRANS code to represent Tamil symbols[3].

To develop sound units for Tamil one has to identify the units for the task. For English, phonemes are used as the sound units. For Tamil language, the English phonemes may not be the best choice. But in order to express the words in terms of pronunciation units, we consider phone-like units with which we can build the pronunciation dictionary for Tamil words

As no phoneme set is defined for Tamil, we identify the sound units for the Tamil language and assign a symbol to each. In this case the symbols are chosen to be the same as those used for English ARBABET phone set for similar sounds [4]. For sound units that are not available in the English set, new symbols are assigned. The 'BEEP' dictionary for English is used to find similar sounds and their appropriate phoneme transcriptions [5]. The lexicon used for the two tasks are listed in Tables 3 and 4.

Hidden Markov models are created for each of the sound units. To create and train the HMMs, the HMM Toolkit is used [6]. From the speech waveform a 39 dimensional parameter vector is derived for each frame of 20msec [1]. The 39 dimensional feature vector consists of 13 melcepstral coefficients, 13 delta cepstral coefficients and 13 delta delta cepstral coefficients. A three state left to right no skip HMM is used. The means and variances of the models are initialised to the global means and variances. These flat start monophones are reestimated using Baum-Welch reestimation on the speech data. A silence model 'sil' is added to make the model robust against impulsive noises in the training data, and to take care of multiple pronunciations.

Speech data for testing is recorded and the parameter vectors are extracted for each 20msec frame. Word level transcriptions of these files are created. The output of the recogniser is compared with the reference transcription, and the results are evaluated. The word and sentence level accuracy for the TV control application are 99.2% and 98.3% respectively. For the banking application the word accuracy is 95% and the sentence accuracy is 28%. The high accuracy for the TV control application is due to the small vocabulary and short sentences. The fall in sentence accuracy for the banking application is due to the large number of words per utterance (20-30 words). But the word accuracy is still reasonably high. A sample of the test utterances are listed in tables 5 and 6.

3. CONCLUSION

In this paper we have shown that it is possible to build a functional system for interacting with a machine in restricted task situations in a local language like Tamil, using the available tools such as the HMM Toolkit. The available symbols for the pronunciation units of English can be adopted for the new language. New symbols are created to take into account the sound units in Tamil not available in the English pronunciation dictionary. The recognition system is speaker dependent, and uses context free monophones for recognition. Speaker independence can be acheived partially by exploring bigram, trigram models with context dependancy, and also by incorporating language constraints into the recognition at higher levels.

4. REFERENCES

[1] L.R.Rabiner and B.H.Juang, *Fundamentals of Speech Recognition.* New Jersy: Prentice Hall, 1993.

[2] Steve Young, Dan Kershaw, Julian Odell, Dave Ollason, Valtcho Valtchev, Phil Woodland, *the htk book version 2.2.* New Jersy: Entropic, 1999.

[3] Description: ITRANS - Indian Language Transliteration Package Version 5.2 , *Source http://www.aczone.com/itrans/.*

[4] John R. Deller Jr, John G. Proakis, John H.L. Hansen, *Discrete-Time Processing of Speech Signals.* New York: Mc Millan Publishing Company, 1993.

[5] Description: Phonemic transcriptions of over 250,000 English words. (British English pronunciations), *Source ftp://svr-ftp.eng.cam.ac.uk/comp.speech/dictionaries/ beep.tar.gz.*

[6] HTK is a toolkit for building Hidden Markov Models, htk ver 2.2, *by Entropic.*

<table>
<tr><th colspan="2">Table 1 Task grammar for the TV application</th></tr>
<tr><th>Variables</th><th>Optional values</th></tr>
<tr><td>$onoff</td><td>A^n | Af;</td></tr>
<tr><td>$command1</td><td>ca^nel;</td></tr>
<tr><td>$command2</td><td>oli;</td></tr>
<tr><td>$option1</td><td>rAj [TIvi] | vijay [TIvi] | Diskavari [ca^nel] |</td></tr>
<tr><td></td><td>aDutta ca^nel | mu^ndiya ca^nel ;</td></tr>
<tr><td>$option2</td><td>aigam [adiga] | kurai [kuraiya];</td></tr>
<tr><td colspan="2">(SENT-START ((TIvi | oli) $onoff | $command1 $option1 | $option1 |</td></tr>
<tr><td></td><td>$command2 $option2 | $option2 $command2) SENT-END)</td></tr>
</table>

<table>
<tr><th colspan="2">Table 2: Task grammar for the banking task</th></tr>
<tr><th>Variables</th><th>Optional values</th></tr>
<tr><td>$digits</td><td>o^nRu | iraNDu | mUNRu | nANgu | aindu | ARu | Ezu | eTTu | o^nbaDu |bUjiyam ;</td></tr>
<tr><td>$account</td><td>esbi akauNT |cEmippu va~Ngi kaNakku eN | kareNT akauNT |naDappu kaNakku eN;</td></tr>
<tr><td>$rdaccount</td><td>ArDi akauNT | todar vaippu kaNakku eN ;</td></tr>
<tr><td>$accountno</td><td>$digits $digits $digits $digits $digits $digits;</td></tr>
<tr><td>$passwd</td><td>telipin eN $digits $digits $digits $digits;</td></tr>
<tr><td>$transaction1</td><td>rokkam varavu peRRukkoLga | kAcOlai varavu peRRukkoLga ;</td></tr>
<tr><td>$transaction11</td><td>kAcOlai peRRu kaNakku kazittu rokkam taRuga |</td></tr>
<tr><td></td><td>todar vaippu eN $accountno ukku mARRuga |</td></tr>
<tr><td></td><td>kazittu ce^n^nai telipO^ns peyaril kAsOlai ezudi taRuga ;</td></tr>
<tr><td>$transaction2</td><td>vaDDi vivaram taRuga | mIdi iRuppu vivaram taRuga;</td></tr>
<tr><td>$transaction3</td><td>esbiai mayilai kiLai ;</td></tr>
<tr><td>$transstart</td><td>tra^nsakshans Arambam ;</td></tr>
<tr><td>$transend</td><td>tra^nsakshans muDivu ;</td></tr>
<tr><td>(SENT-START</td><td>($transstart ($account $accountno $passwd (rUbai <$digits></td></tr>
<tr><td></td><td>$transaction1 |rUbai <$digits> $transaction11 | $transaction2 | $transaction3</td></tr>
<tr><td></td><td>$accountno ukku rUbai <$digits> mARRuga) $transend) | ($transstart $rdaccount</td></tr>
<tr><td></td><td>$accountno $passwd (rUbai <$digits> $transaction1 | $transaction2)) $transend</td></tr>
<tr><td></td><td>) SENT-END)</td></tr>
</table>

Table 3 Task vocabulary for theTV application (ITRANS code and Tamil equivalent)

A^n	ஆன்	Af	ஆப்	Kurai	குரை	rAj	ராஜ்
adiga	அதிக	ca^n	சன்	Kuraiya	குரைய	TIvi	டிவி
adigam	அதிகம்	ca^nel	சனெல்	mu^ndiya	முந்திய	vijay	விஜய்
aDutta	அடுத்த	Diskavari	டிஸ்கவரி	Oli	ஒலி		

Table 4: Task vocabulary for the banking task (ITRANS code and Tamil equivalent)

ARu	ஆறு	kAcOlai	காசோலை	Peyaril	பெயரில்
Akkit	ஆக்கித்	kAsOlai	காசோலை	RUbai	ரூபாய்
ArDi	ஆர்டீ	kaNakku	கணக்கு	Rokkam	ரொக்கம்
Arambam	ஆரம்பம்	kareNT	கரெண்ட்	TaRuga	தறுக
DepAsit	டேபாசிட்	kazittu	குழித்து	Taruga	தருக
Ezu	ஏழு	kiLai	கிளை	Tavanai	தவணை
aindu	ஐந்து	lO^n	லோன்	telipO^ns	டெலிபோன்ஸ்
akauNT	அகௌண்ட்	mARRuga	மாற்றுக	Telipin	டெலிபின்
bUjiyam	புஜியம்	mIdi	மீதி	Todar	தொடர்
cEmippu	சேமிப்பு	mUNRu	முண்று	tra^nsakshans	டிரன்ஸக்ஷன்ஸ்
cITTu	சீட்டு	mayilai	மயிலை	ukkA^na	உக்கான
ce^n^nai	சென்னை	muDivu	முடிவு	UkkAna	உக்கான
eN	எண்	nANgu	நாண்கு	Ukku	உக்கு
eTTu	எட்டு	naDappu	நடப்பு	VaDDi	வட்டி
esbi	எஸ்பி	nagai	நகை	VaRuda	வருட
esbiai	எஸ்பிஐ	nilai	நிலை	Vaippu	வைப்பு
ezudi	எழுதி	o^nRu	ஒன்று	Varavu	வரவு
fiksed	ஃபிக்செட்	o^nbaDu	ஒன்பது	va~Ngi	வங்கி
iRuppu	இருப்பு	peRRu	பேற்று	Vivaram	விவரம்
iraNDu	இரண்டு	peRRukkoLga	பேற்றுக்கொள்க		

Table 5 Sample test utterance for the TV application

1	aDutta ca^nel	5	oli adigam
2	mu^ndiya ca^nel	6	TIvi Af
3	oli kurai	7	TIvi A^n

Table 6 Sample test utterances for the banking application

1	SENT-START tra^nsakshans Arambam cEmippu va~Ngi kaNakku eN bUjiyam Ezu iraNDu iraNDu Ezu Ezu telipin eN nANgu o^nRu nANgu o^nRu rUbai eTTu eTTu eTTu Ezu rokkam varavu peRRukkoLga tra^nsakshans muDivu SENT-END
2	SENT-START tra^nsakshans Arambam ArDi akauNT ARu eTTu mUNRu o^nbaDu bUjiyam o^nbaDu telipin eN o^nRu mUNRu mUNRu mUNRu mIdi iRuppu vivaram taRuga tra^nsakshans muDivu SENT-END
3	SENT-START tra^nsakshans Arambam naDappu kaNakku eN eTTu iraNDu eTTu o^nbaDu o^nRu iraNDu telipin eN Ezu Ezu ARu bUjiyam mIdi iRuppu vivaram taRuga tra^nsakshans muDivu SENT-END