OPEN ACCESS Information Technology Journal

ISSN 1812-5638
DOI: 10.3923/itj.2017.

Research Article

Context-dependent Syllable Modeling of Sentence-based Semi-continuous Speech Recognition for the Tamil Language

[1]Ibralebbe Mohamed Kalith, [2]David Asirvatham and [3]Ismail Raisal

[1]Department of Mathematical Sciences, Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sammanthurai, Srilanka
[2]School of Computing and IT, Taylor's University, Subang Jaya, Selangor, Malaysia
[3]Department of Management, Faculty of Management and Commerce, South Eastern University of Sri Lanka, Oluvil, Srilanka

Abstract

Background and Objective: Automatic speech recognition transcribes acoustic signal into strings of words for a given language. Speech recognition applications in native languages will enable information accessible for illiterate and disabled user in the society. In this research, the focus was on improving automatic speech recognition of the Tamil language. The development of a large-vocabulary continuous speech recognition system for Tamil language, which requires an acoustic model to be trained on a large vocabulary corpus. **Methodology:** To address the challenges, a modeling efficient sub-word units were recommended and designed a consonant-vowel six-segment (CVS-6) algorithm for syllabification of a Tamil text corpus and experimentally investigated its speech recognition accuracy. A specific database was constructed using 120 sentences of semi-continuous speech, comprising 561 words and 436 unique syllables. **Results:** The syllable-based model achieved a mean recognition rate of 81.41% (standard deviation, 6.94%) compared with the 69.87% (standard deviation, 4.11%) achieved by a phoneme-based model. The word error count for complex words was 25% by the syllable-based model compared with 54.96% by the phoneme-based model, a reduction of 30%. **Conclusion:** Syllable-based model using consonant-vowel six-segment algorithm is good choice and can be used to sub-word modeling of large vocabulary continue speech in Tamil language.

Key words: HMM, syllable-based, phoneme-based, semi-continuous, Tamil language, consonant-vowel six-segment algorithm

Citation: Ibralebbe Mohamed Kalith, David Asirvatham and Ismail Raisal, 2017. Context-dependent syllable modeling of sentence-based semi-continuous speech recognition for the Tamil language. Inform. Technol. J., CC: CC-CC.

Corresponding Author: I.M. Kalith, Department of Mathematical Sciences, Faculty of Applied Science, University Park, South Eastern University of Sri Lanka, Sammanthurai, Sri Lanka Tel: +94 776022868

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Information access is a key factor in a globalizing world. Human-computer interaction technology has made information more accessible. People interact with computers in several ways, the computer interface must be designed to facilitate this interaction. The most popular interface today is the graphical user interface (GUI), adopted by desktop applications, internet browsers, mobile computers and mobile phones. Voice user interfaces (VUI) are used for speech recognition (SR) and the emerging multi-modal and gestalt GUIs allow users to engage with embodied agents[1]. The SR is a sub-field of computational linguistics, which develops methodologies and technologies to enable human speech to be recognized and translated to text by a machine. This is also known as automatic speech recognition (ASR).

The SR technology uses a fundamental acoustic unit called sub-word. The SR research involves the identification of possible sub-words of a given language. However, major problems remain in the modeling of these sub-word structures. In spoken language, sub-word units usually represented by its phonemes and syllables. Modeling of a sub-word unit from phonemes and syllables is therefore a key technology in SR. The word is often regarded as the basic unit of language, because it has a defined boundary in speech but in continuous speech, word boundaries are blurred. Tamil is an agglutinative language[2], in which complex words are formed by chaining morpho-phonological strings. In this ASR research study, Tamil language sub-word units are modeled, parameterized and used for training. Modeling of phonemes and syllables of the Tamil language was performed using sentence-based semi-continuous speech recognition (SB-SCSR) and the impact of sub-word modeling on the rate of recognition was investigated.

Many studies in many languages have proposed high-performance ASR systems based on sub-word units such as phonemes and syllables[3]. The performance of these approaches has been compared with that of context-independent (CI) and context-dependent (CD) acoustic modeling. Thangarajan[4] developed word-based CI acoustic models for 371 Tamil words and triphone-based CD acoustic models for 1700 words using CMU (Carnegie Mellon University, USA) Sphinx tools. It is concluded that when applied to Tamil ASR, CI word modeling yielded reasonable accuracy for a small vocabulary but the triphone-based approach was more suitable for medium and large vocabularies. Saraswathi and Geetha[5] proposed a phoneme recognition system for Tamil words, in which the performance was enhanced by using language models (LMs) at the recognition phase. Speech signals were segmented using LMs and recognition was achieved using a similarity measure, based on the acoustic characteristics of the phoneme signal in speech. The errors in the recognized phoneme sequence were detected and corrected using an integrated variable length phoneme model and inter-word hybrid language model. Weerasinghe *et al.*[6] investigated the syllable structure of Sinhala language. It was proposed that an algorithm for identifying syllables in Sinhala words by identifying sets of rules. The algorithm was tested on 30,000 distinct words from a corpus and the results were compared with manual identification of syllables of the same words. The algorithm performed with 99.95% accuracy. Ganapathiraju *et al.*[7] built the first successful robust LVCSR system that applied a syllable-level acoustic unit to spontaneous telephone speech. It was demonstrated that the superiority of their system to existing triphone systems. However, the system proved deficient in integrating the syllable and phoneme models as a mixed-word entry. Mixing models of different lengths and contexts produced only marginal improvements. Lakshmi and Hema[8] proposed a syllable-based continuous speech recognizer using group delay based two-level segmentation to mine syllables from a Tamil speech corpus. This did not require annotated transcribed training data. It was concluded that the novel recognizer produced better results than a conventional hidden Markov model (HMM) based continuous recognizer, due to the reduction of complexity in the search space.

Tamil language is a member of the Dravidian language family. It has been spoken by about 77 million people globally, making it the 15th most widely spoken language[9]. It has been known as a lingua franca in South India, Sri Lanka, Malaysia and Singapore and has been widely spoken in the Northern and Eastern parts of Sri Lanka, where it has been recognized as an official language. There was some dialectical variation in Sri Lanka, with different speech patterns in the upcountry area of Nuwara Eliya, the northern areas of Jaffna and in the eastern parts.

Tamil scholars and linguists classify the history of Tamil language and literature into 5 periods: The Sangam period (300 BC-300 AD), Association time (300 AD-700 AD), Bhakti literature period (700 AD-1200 AD), Middle period (1200 AD-1800 AD) and modern period (1800 AD-present)[10]. The phoneme was a fundamental unit of speech. Tamil phonology uses 12 vowels (V), 18 consonants (C), 216 compounds (VC) and a secondary character called the āytam. There has been 247 letters in the standard Tamil alphabet and the Tamil language has almost 100 sounds[10]. Special features of Tamil language phonology include retroflex consonants and multiple rhotics[11]. Tamil language uses 12 vowel sounds[10]. Their places of articulation were shown in Table 1. There were

Table 1: Tamil phoneme representation in IPA and ARPABET Format

ARPABET	IPA	Tamil
AH	ʌ	அ – அப்பா
AA	a:	ஆ – ஆணி
IH	ɪ	இ – இந்தி
IY	ɪ̈	ஈ – ஈ
UH	ʊ	உ – உலகம்
UW	u:	ஊ – ஊர்தி
EH	ɛ	எ – எலி
EY	əɪ	ஏ – ஏணி
AY	aɪ	ஐ – ஐம்பது
AO	ɔ	ஒ – ஒரு
OH	ɔ:	ஓ – ஓடு
AW	aʊ	ஔ – ஔடதம்
K	k	க – கலை
G	g	க – அங்கே
HH	h	க – பகல்
NG	ŋ	
CH	tʃ	ச – பச்சை
S	s	ச – சட்டை
J	j	ச – பஞ்சு
NC	ɲ	
T	t	ட – பாட்டு
D	ɖ	ட – நாடு
NX	n	ண – கண்
TH	t̪	த – பத்து
DH	d̪	த – அது
NH	ɳ	ந – பந்து
P	p	ப – பத்து
B	b	ப – கோபம்
M	m	
Y	j	ய – கொய்யா
RR	R	
L	l	ல – பல்
V	v	வ – செவ்வாய்
Z	L	ழ – தமிழ்
LL	ɭ	ள – கடவுள்
R	r	ற – கறை
N	N	ன – நான்

18 basic consonants in the Tamil language. Taking account of these allophones[4], Tamil has approximately 25 distinct consonant phonemes. The phonology of Tamil was summarized in Table 1.

Tamil language is an agglutinative language with complex inflectional morphology and morpho-phonology, in which words were formed by combining a sequence of phonemes. That was, a word may consist of many phonemes. This makes recognizing the boundary of a complex word challenging in ASR. When applying ASR to Tamil language, therefore, a large dataset was needed to train the acoustic model. One way of addressing this was the application of efficient sub-word modeling. In this study, the effectiveness of sub-word modeling, using phonemes and syllables, in recognizing complex Tamil words was investigated. These complex words were frequent in large Tamil corpora.

Agglutinative inflectional morphology: Words were combined the same root word with different prefixes and suffixes[2]. Example: The verbs in Tamil form inflectional variations (Table 2). Two meaningful words were combine to form new words called Morpho-phonology, in which words were modified by deletion, insertion, or substitution of characters at the word boundary to form a new word, was widely used in Tamil[2] (Fig. 1).

The main objective of this study was to develop an ASR model that can recognize continuous speech in spoken Tamil language and specifically to investigate that how proposed syllable-based sub-word modeling can contribute

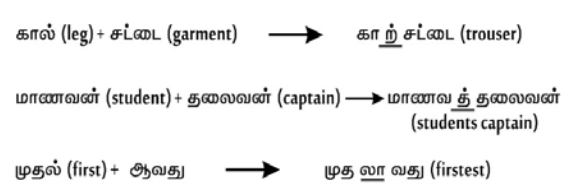

Fig. 1: Examples of morpho-phonological word formation

Table 2: Example of agglutinative (inflectional variations) word formation for tense markings

English			Tamil		
			Verb with tenses		
Noun	Verb	Noun	Past	Present	Future
I	Came	p	வந்தேன்	வருகிறேன்	வருவேன்
You	Came	நீ	வந்தாய்	வருகிறாய்	வருவாய்
They	Came	அவர்கள்	வந்தனர்	வருகிறார்கள்	வருவார்கள்
He	Came	அவன்	வந்தான்	வருகிறான்	வருவான்
She	Came	அவள்	வந்தாள்	வருகிறாள்	வருவாள்

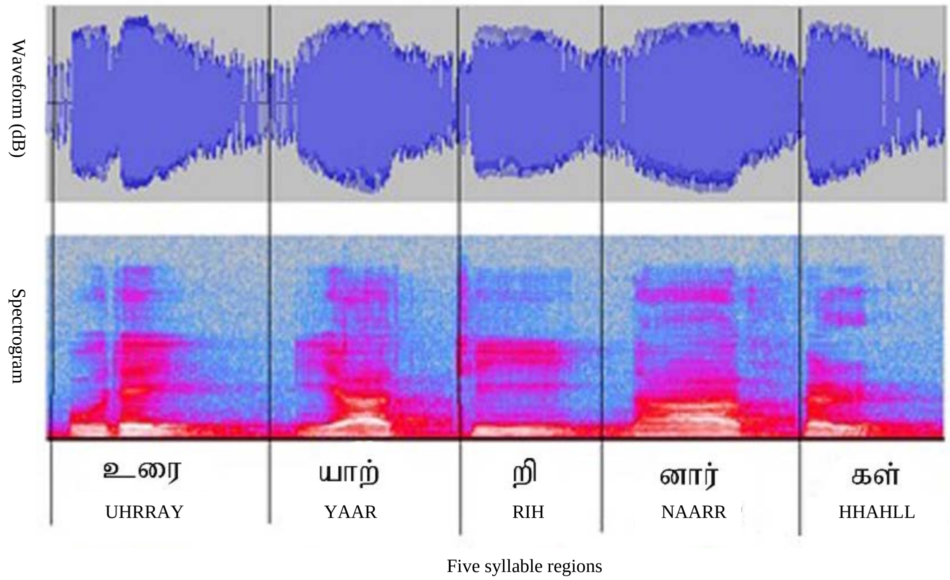

Fig. 2: Waveform and spectrogram analysis of complex word showing 5 syllable regions

to the recognition of complex Tamil language words. Given the resource limitations, it was unable to build a full large-vocabulary continuous speech recognition (LVCSR) system. Instead of it, a medium vocabulary size SB-SCSR model for recognition of sentences in continuous Tamil language speech was constructed.

This study applied manual segmentation and annotation of semi-continuous sentences by analyzing the short-term energy spectra of voice signals, as shown in Fig. 2. By analyzing the formant frequency of the spectra, syllable boundaries could be identified for manual labeling.

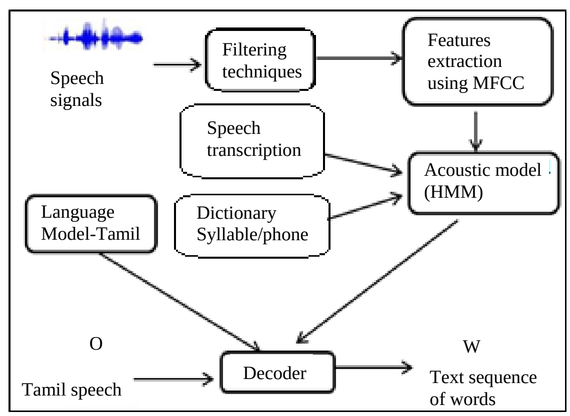

Fig. 3: Automatic speech recognition system for Tamil

MATERIALS AND METHODS

Tamil ASR model: The architecture of the ASR system for Tamil language speech used throughout the experiments was shown in Fig. 3. This HMM-based system was built using CMU Sphinx tools, a portable tool kit for building and manipulating HMMs[12] widely used in SR research. The data were sampled at 16 kHz and frame features were extracted to reduce the amount of information in the input signal. In total, 39 mel-frequency cepstral coefficients were extracted at a frame rate of 10 ms, using a 25 ms Hamming window. Satori *et al.*[13] and Abushariah *et al.*[14] investigated automatic speech recognition for Arabic language by implementing CMU Sphinx speech recognizing tools. The acoustic model were trained and tested with sphinx-4 decoder.

Text segmentation into phonemes: In all languages, words were combined to form sentences and sentences are combined to form continuous speech. In SR systems, the most commonly-used sub-word unit was the phoneme. In this research, the 37 phonemes of Tamil were used (Table 1) to represent the entire speech database in SB-SCSR. Phonemes were typically represented by triphone modeling, as shown in Fig 4. The HMM applied Gaussian mixture observation densities to account for acoustic variation between different instances of the same unit.

The same grapheme may be pronounced differently in different words, as articulation was influenced by context an example in which the pronunciation of a grapheme changes by its relation with adjacent phones was shown in Table 1.

உரையாற்றினார்கள்	உ	ரை	யா	ற்	றி	னா	ர்	க	ள்	9 Letters
Phoneme	UH	RRAY	YAA	R	RIH	NAA	RR	HHAH	LL	14 Phonemes
Tri-Phones	{UHRRAY}*, {RRAYY}, {AYYAA}, {YAAR}, {AARR}, {RR IH}, {RIHN}, {IHNAA}, {NAARR}, {AARRHH}, {RRHHAH} and {HHAHLL}*									12 tri-phones
Syllables (CVS-6 algorithm)	உரை UHRRAY		யாற் YAAR		றி RIH	னார் NAARR		கள் HHAHLL		5 syllables

Fig. 4: Triphone clustering and syllabification

```
Count word length = n
Position of letter = P
For i = 1 to n do
•   If Pi = V AND Pi+1 = CV, Then
      Check Pi+2,
        If Pi+2 = C, Then
           Pi+1, Pi+2 is a Syllable [CV.C]
        Else
           Pi, Pi+1 is a Syllable [V.CV]
Else if Pi = V AND Pi+1 = C, Then
           Pi, Pi+1 is a Syllable [V.C]
Else
      Pi is a Syllable [V]

•   If Pi = CV AND Pi+1 = C
Check Pi+2,
        If Pi+2 = C, then
           Pi, Pi+1, Pi+2 is a Syllable [CV.C.C]
      Else
           Pi, Pi+1 is a Syllable [CV.C]
Else
           Pi is a Syllable [CV]
```

Fig. 5: Proposed CVS-6 syllabification algorithm

When the grapheme was pronounced in the middle of a word, such as , it has the IPA equivalent /h/. When pronounced at the start of the word, such as , it has the equivalent /k/ and at the end of a word, such as , the equivalent is /g/. Thus, the acoustic variability in basic phonetic units due to context is large and not well studied for many languages. Bahl *et al.*[15], reported that, in such contexts, a word-based model performed significantly better than a phoneme-based model in dynamic time warping (DTW). Therefore, a phoneme-based model may not always be optimal when addressing highly context-sensitive language.

Text segmentation into syllables: A syllable was a sub-word unit that may be smaller than a word or larger than a single-phoneme. A syllable typically spans several phonemes, making them easier to recognize in a search[16,17]. However, defining a syllable is problematic in the Tamil language, as

Table 3: Six types of syllable produced by CVS-6 syllabification algorithm with suitable example of Tamil language syllables

Syllable	Example
V	அ, ஆ, இ, ஈ, உ, ஊ......
CV	க, கா, த, தா, ல, லை, சை, தொ
V.CV	ஆகா, இலை, ஊசி,
CV.C	தட், டிப்,றிப், வன், கண்,
V.C	ஆல், இன், உன், அப்,
CV.C.C	வீழ்ந், வாழ்க், கிர்ந்,

different languages have different rules of syllabification. One simple approach was a consonant-vowel six-segment (CVS-6) algorithm for syllabification. In the short energy spectrum analysis (Fig. 2), formant value of the energy spectrum is shown to be almost co-extensive with a syllable segment represented by the CVS-6 algorithm.

Algorithm for CVS-6 syllabification of Tamil words: This algorithm segments text using consonant and vowel identifiers. As the syllable was categorized into six segments as shown in the Table 3, hence, named this consonant-vowel six-segment (CVS-6) algorithm. The CVS-6 algorithm has been unique as it uses Tamil linguistic rules to segment syllables. Lakshmi and Hema[8] have proposed an algorithm that can segment syllables in transliterated or Romanized Tamil text, this algorithm forms lot of ambi-syllabic consonants but CVS-6 algorithm minimized ambi-syllabic consonants also this algorithm was to segment the syllables as written in Tamil, as shown in Fig. 5.

RESULTS AND DISCUSSION

To test the proposed syllable-based sub-word model recognition accuracy, experiments are conducted using open source CMU Sphinx speech recognition toolkit. In this research two experiments were designed in order to evaluate the performance of sub-word modeling of speech recognition. The first was phoneme-based modeling of SB-SCSR and the second was syllable-based modeling of SB-SCSR.

Database: The sentence-based semi-continuous model was trained for medium vocabularies of Tamil speech and a sentence-level corpus to allow the acoustic models to recognize sentences. To test our objective of a sub-word effect in the recognition of Tamil speech, a special database was constructed. The data were collected and analyzed using the following steps.

Table 4: ASR model training and decoding with semi-continuous sentences (phoneme-based and syllable-based)

Decode attempt	Number of sentences for training	Number of sentences decoding	Number of words in decoding	Phoneme-based Correct word	Error	Accuracy (%)	Syllable-based Correct word	Error	Accuracy (%)
Test 1	20	4	17	12	5	70.59	14	3	82.35
Test 2	40	8	30	19	11	63.33	24	6	80.00
Test 3	60	12	44	30	14	68.18	30	14	68.18
Test 4	80	16	52	36	16	69.23	44	8	84.62
Test 5	100	20	76	55	21	72.37	65	11	85.53
Test 6	120	24	98	74	24	75.51	84	14	85.71
Mean						69.87			81.07
SD						±4.11			±6.68

Table 5: Speech files parameters

Parameters	Value
File type	mswav
File extension	wav
Sampling rate	16 KHz
Depth	16 bits
Mono/Stereo	Mono
Feature file extension	mfc

Step 1: Around 120 Tamil sentences were selected for training the acoustic speech model. These were collected from primary-level school textbooks, newspapers, articles, novels and magazines. The collected sentences were then organized into a batch system from Test 1-6 (Table 4) for training and testing

Step 2: The collected sentences were uttered by a native speaker and recorded. The following conditions were applied

- The speaker were aged from 20-40 years
- All were speakers of Sri Lankan Tamil
- Each word was articulated clearly

Step 3: Recording was carried out under the conditions shown in Table 5. Audacity software was used for recording and editing, as it is a powerful and sophisticated tool allowing segmentation of the speech signal and noise removal while maintaining speech parameters

Step 4: Speaker recorded 120 sentences (561 words). Each sentence was spoken thrice and the clearest version was selected. This semi-continuous speech data was used in experiments 1 and 2. The acoustic model was trained using this corpus, phoneme-based and syllable-based sub-word mappings were compared to assess their suitability for Tamil ASR system development

Experiment 1 (phoneme-based modeling of SB-SCSR): In this experiment, the selected sentences, annotated for segmentation using a phoneme-based sub-word model were used. Training was done using a CD or triphone-based semi-continuous modeling. The training was conducted in batch mode and each batch of training data was tested with the recorded test data, as specified in Table 4. The experimental results are also given in Table 4. The scatter diagram in Fig. 6 shows the overall performance of the phoneme-based sub-word model. The mean recognition rate was 69.87%, with a standard deviation of 4.11%. A strong correlation can be observed between an increase in the size of the training database and an increase in the recognition rate. It is assumed that this reflects the distribution of complex words in the test data set. The total word error (WE) count and the number and percentage of total WEs involving complex words (agglutinative) as shown in Table 6. The mean WEs was 54.96% for phoneme-based.

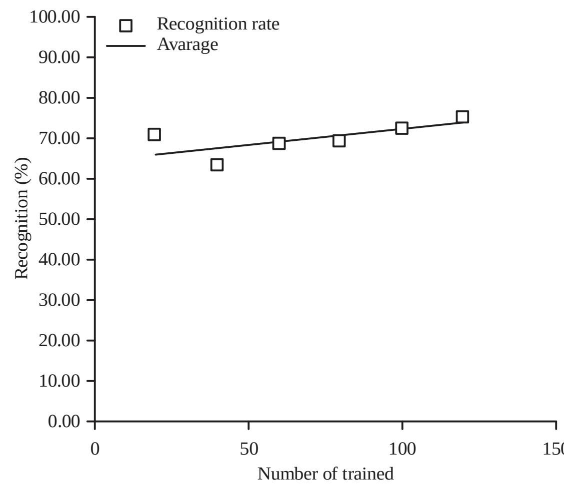

Fig. 6: Scatter diagram of recognition rate for phoneme-based sub-word modeling of semi-continuous speech

Experiment 2 (syllable-based modeling SB-SCSR): The second experiment addressed the recognition of agglutinative and morpho-phonologically complex words in Tamil. The model was trained and tested and the results were summarized in Table 4. This experiment used the same

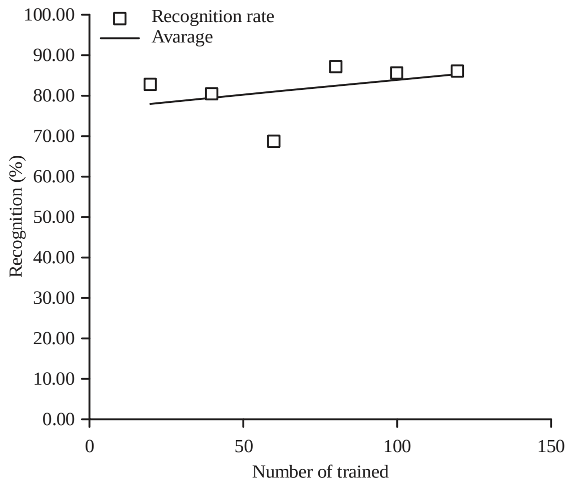

Fig. 7: Scatter diagram of recognition rate for syllable-based sub-word modeling of SB-SCSR

Table 6: Word error (WE) counts for complex words (phoneme-based)

Decode attempt	Word error count	WE in complex word count	Complex WE (%)
Test 1	5	3	60.00
Test 2	11	6	54.55
Test 3	14	8	57.14
Test 4	16	9	56.25
Test 5	21	10	47.62
Test 6	24	13	54.17
Mean			54.96

Table 7: Word error (WE) counts for complex words (syllable-based)

Decode attempt	Word error count	WE in complex word count	Complex WE (%)
Test 1	3	0	0.00
Test 2	6	2	33.33
Test 3	14	2	14.29
Test 4	8	3	37.50
Test 5	11	4	36.36
Test 6	14	4	28.57
Mean			25.01

selected sentences and annotated speech data as experiment 1 but segmented according to the CVS-6 syllable-based sub-word model. Training was conducted in CD mode based on semi-continuous modeling. The training was conducted in batch mode and each batch of training data was tested with the recorded test data shown in Table 4. The experimental results were also shown in Table 4. The scatter diagram in Fig. 7 shows the overall results for the syllable-based sub-word modeling of semi-continuous speech. The mean recognition rate was 81.07%, with a standard deviation of 6.94%. As can be seen from Fig. 7, a strong correlation was again observed between an increase in the size of the training database and an increase in the recognition rate. The total WEs and the number and percentage of total WEs involving complex words as shown in Table 7. The mean WEs was 25% for syllable-based.

Currently, a few studies were focused on developing syllable segmentation algorithm in segmenting text and speech. The syllable segmented from a text using syllabification algorithm and characteristic part was identified from the formant value of spectrum in the speech signals. Other than the algorithm marked the boundary of the syllable in the speech spectrum, these syllables were resemble with syllable in the text. This study was followed the first method of syllable segmentation. Natarajan and Jothilakshmi[18] proposed an algorithm for segmentation of the syllable from the continuous speech signal. After removing silence from speech signals the speech segment further processed using linear predictive coding (LPC) to extract the formant frequencies. These format frequencies were used to mark the boundary of the syllables. The proposed algorithm was produced accuracy of 89% identifying syllable boundary when it's comparable with hand labeled syllable boundaries. Venkataramana *et al.*[19] developed a Telugu speech recognition system using HMM, achieved 81% word accuracy in speaker independent. In this study, the Telugu morpheme was used as a fundamental sub-word unit for acoustic modeling. Morpheme is a smaller unit than syllable, but bigger than phoneme and the phonetic base of Telugu language is different from Tamil language. Akila and Chandra[20] used time normalization and rate of speech to performance enhancement of syllable-based Tamil speech recognition. Speech rate dependent syllable-based Tamil SRs. Format value of characters used to segment the syllable. The database used in this study was 200 syllables and 4 speakers for training. The baseline 70% accuracy and the proposed model shown 74% accuracy. This study was maintained the rate of speech in average and the syllable boundaries were clearly identified by a slow rate of speech. Geetha and Chandra[21] proposed a monosyllable isolated word recognition for Tamil language using continuous density HMM. Fourty two monosyllable Tamil words, used single female speaker. Monosyllable as sub-word the result shows 5 state model with four mixture component is enough to implement monosyllable or character based Tamil IWR. Hlaing and Mikami[22] proposed a new method for Myanmar syllabification which deployed formal grammar and un-weighted finite state transducers (FST). The algorithm focuses on orthographic ways of syllabification for the input tests encoded in unicode and the algorithm perform 99.93% accuracy on syllabification. Thangarajan and Natarajan[23] note that the pronunciation of Tamil speech is based on prosodic syllables. The algorithm was developed to segment the words into prosodic syllables and the authors trained the acoustic model with 1398 unique syllables in the CI mode. The model was tested using 45 sentences with 617 words and the results were obtained with 80% accuracy. The results show that accuracy decreased by 10% compared with the baseline triphone model. In contrast, the acoustic model in this study was trained with 436 distinct syllables in the CD mode. The results showed that the speech recognition accuracy of the syllable-based model was 11% better than that of the baseline phoneme-based model and gave a reduction of 30% in WE rates.

CONCLUSION

This study compared two approaches to sub-word modelling for speech recognition in Tamil language: Phoneme-based and syllable-based modelling. The conventional phoneme-based model were faced challenges of recognizing complex words in continuous speech. The study proposed a new CVS-6 algorithm for syllable segmentation of the Tamil text corpus. From the experimental results the syllable-based sub-word model was shown a 30% reduction in complex word error compare with baseline phoneme-based model. Overall, the proposed syllable-based model produced an improvement of 11% in recognition. The result clearly indicates that the syllable-based sub-word modeling performs better for sentence-based semi-continue and large vocabulary continues speech in Tamil language.

Further study was needed to improve ASR of Tamil. There were four key tasks: The development of Tamil speech corpora, improved language modeling, the development of acoustic models for speech recognizers and the applications of those speech recognizers.

SIGNIFICANCE STATEMENTS

This study explored the possibility of using syllable-based sub-word modelling for the recognition of agglutinative Tamil words in sentence-based semi-continuous speech. Such approaches are required for the development of speech recognition systems for highly agglutinative languages. A novel CVS-6 algorithm was developed to support the syllabification.

REFERENCES

1. Rabiner, L.R., J.G. Wilpon and F.K. Soong, 1988. High performance connected digit recognition using hidden Markov models. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, April 19-24, 1988, IEEE., New York, pp: 119-122.
2. Thangarajan, R., 2012. Speech Recognition for Agglutinative Languages. In: Modern Speech Recognition Approaches with Case Studies, Ramakrishnan, S. (Ed.)., InTech, Croatia.
3. Anusuya, M.A. and S.K. Katti, 2009. Speech recognition by machine: A review. Int. J. Comput. Sci. Inform. Secur., 6: 181-205.
4. Thangarajan, R., A.M. Natarajan and M. Selvam, 2008. Word and triphone based approaches in continuous speech recognition for Tamil language. WSEAS. Trans. Signal Process., 4: 76-86.
5. Saraswathi, S. and T.V. Geetha, 2007. Comparison of performance of enhanced morpheme-based language model with different word-based language models for improving the performance of Tamil speech recognition system. ACM Trans. Asian Language Inform. Process., Vol. 6. 10.1145/1290002.1290003.
6. Weerasinghe, R., A. Wasala and K. Gamage, 2005. A Rule-Based Syllabification Algorithm for Sinhala. In: Natural Language Processing, Dale, R., K.F. Wong, J. Su and O.Y. Kwong (Eds.)., Vol. 3615, Springer, Berlin, Heidelberg, pp: 438-449.
7. Ganapathiraju, A., J. Hamaker, J. Picone, M. Ordowski and G.R. Doddington, 2001. Syllable-based large vocabulary continuous speech recognition. IEEE Trans. Speech Audio Process., 9: 358-366.
8. Lakshmi, A. and A.M. Hema, 2006. A syllable based continuous speech recognizer for Tamil. Proceedings of the 9th International Conference on Spoken Language Processing, September 17-21, 2006, Pittsburgh, Pennsylvania, pp: 1878-1881.
9. Schiffman, H.F., 2006. A Reference Grammar of Spoken Tamil. Cambridge University Press, Cambridge.
10. Keane, E., 2004. Tamil. J. Int. Phonetic Ass., 34: 111-116.
11. Kalith, I.M. and S. Thelijjgoda, 2011. Automatic speech recognition system for Tamil language. Proceedings of the 29th National Conference on Exploiting ICT Trends for Collaborative Development, August 29-31, 2011, Colombo, pp: 81-85.
12. Lamere, P., P. Kwok, W. Walker, E. Gouvea, R. Singh, B. Raj and P. Wolf, 2003. Design of the CMU sphinx-4 decoder. Proceedings of the 8th European Conference on Speech Communication and Technology, September 1-4, 2003, Geneve, Switzerland, pp: 1181-1184.
13. Satori, H., H. Hiyassat, M. Harti and N. Chenfour, 2009. Investigation arabic speech recognition using CMU sphinx system. Int. Arab J. Inform. Technol., 6: 186-190.
14. Abushariah, M., R. Ainon, R. Zainuddin, M. Elshafei and O.O. Khalifa, 2012. Arabic speaker-independent continuous automatic speech recognition based on a phonetically rich and balanced speech corpus. Int. Arab J. Inform. Technol., 9: 84-93.

15. Bahl, L.R., P.F. Brown, P.V. de Souza and M.A. Picheny, 1988. Acoustic Markov models used in the Tangora speech recognition system. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, April 11-14, 1988, USA., pp: 497-500.
16. Fujimura, O., 1975. Syllable as a unit of speech recognition. IEEE Trans. Acoust. Speech Signal Proces., 23: 82-87.
17. Karpagavalli, S. and E. Chandra, 2016. A review on sub-word unit modeling in automatic speech recognition. IOSR J. VLSI Signal Process., 6: 77-84.
18. Natarajan, V.A. and S. Jothilakshmi, 2015. Segmentation of continuous Tamil speech into syllable like units. Indian J. Sci. Technol., Vol. 8. 10.17485/ijst/2015/v8i17/61362.
19. Venkataramana, K., P. Jayaprakash and P.S.N. Babu, 2015. Developing Telugu speech recognition system using Sphinx-4. Int. J. Innovat. Res. Sci. Eng. Technol., 4: 41-48.
20. Akila, A. and E. Chandra, 2014. Performance enhancement of syllable based Tamil speech recognition system using time normalization and rate of speech. CSI Trans. ICT., 2: 77-84.
21. Geetha, K. and E. Chandra, 2015. Monosyllable isolated word recognition for Tamil language using continuous density hidden Markov model. Proceedings of the IEEE International Conference on Electrical, Computer and Communication Technologies, March 5-7, 2015, Coimbatore, India, pp: 1-6.
22. Hlaing, T.H. and Y. Mikami, 2014. Automatic syllable segmentation of Myanmar texts using finite state transducer. Int. J. Adv. ICT Emerg. Regions, Vol. 6.
23. Thangarajan, R. and A.M. Natarajan, 2008. Syllable based continuous speech recognition for Tamil. S. Asian Language Rev., 18: 72-85.