ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Tamil Text to Emotional Speech Conversion with UNL for Translation

M.Aishwarya[1], N.Priya[2]

M.Tech, Department of C.S.E, Bharath University, Chennai, Tamil Nadu, India[1]

Research Scholar, Dept. of C.S.E., Bharath University, Chennai, Tamil Nadu, India [2]

ABSTRACT: Natural Language Processing aims at translating an input Tamil sentence into an equivalent spoke Tamil translation of the sentence. Natural Language processing deals with the interaction between the computers and human. The translation was done by universal networking language (UNL) is employed for the humanized emotional voice as the output. UNL language is independent it is capable of storing information about the original language from which the text has delivered. Once the text has been translated with the help of the UNL formalism, the emotional voice is synthesized from the neutral robotic voice by modifying certain key parameters that include fundamental frequency, intensity, and duration. Loan words are used better pronunciation. Six recordings comprising of outputs expressing angry, happy and sad emotions were played and the corresponded responses were recorded.

KEYWORDS: Natural Language Processing, Universal Networking Language, Loan Words

I.INTRODUCTION

As UNL is language-independent, it is capable of storing information in a way that is independent of the original language from which the text was derived. This generic way of representation makes UNL very efficient for translation and so it has been employed in our project. Translation has been in existence ever since written literature started gaining traction. In modern times, the preponderance of literature has made it mandatory to automate the process of translation which would otherwise be tedious if done manually. Thus various machine translation approaches were developed over the past decades[3]. Communication in human languages is embedded in context and so the major challenge faced by machine translation involves the. Identification of the context in which words are uttered. If the context has been properly identified, then the translation can be accomplished accurately. Some of the major approaches proposed for machine translation include Rule Based, Phrase-Based, Context-Based, Example-Based and Hybrid approach.

II.LITERATURE SURVEY

Synthesized speech can be created by concatenating part of recorded speech which is stored in a database. The mainly significant qualities of a speech synthesis system are naturalness and Intelligibility [1]. Naturalness expresses how intimately the output sounds like human speech, whereas intelligibility is the easiness with which the output is understood. The main function of text-to-speech (TTS) system is to convert an arbitrary text to a spoken Waveform. TTS is one of the major applications of NLP. TTS Synthesizer is a computer based system that should be understand any text clearly whether it was establish in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. Single text-tospeech (TTS) system for Indian languages Attribution CC BY Susan r. Hertz a primary factor in determining the choice of synthesis strategy is the rule-writer's linguistic convictions. For example, the proponents of an approach based on demisyllables claim that many of the influences of adjacent sounds on each other are automatically present in the demisyllable[14]. Even more challenging is handling phenomena that cannot easily be captured in rules Shen Zhang he proposes an approach on emotional audiovisual speech synthesis. Our approach primarily consists of three steps: first we take the text and the target PAD values as input, and employ text-to-speech (TTS) engine to generate neutral speeches. The prosodic word boundaries are automatically predicted by the text analysis module of TTS engine. In our TTS system, maximum entropy (ME) model is used for prosodic word

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

boundary prediction. The expression of human emotion only has gained special attention recently in both audio speech synthesis and talking face animation. To synthesize the dynamic facial expression in continuous speech, the speech acoustic features (e.g., pitch) are taken as important clues to modulate the facial expression on sentence level. 4. System Design and Implementation

III. PROPOSED SYSTEM

A.UNL Enconverter

The process of converting Source text to UNL Graph is called En-conversion. UNL Graph is composed of nodes called as Universal Words (UW) and relations between them.Also, each UW has one or more attributes attached to them.The output from a morphological analyzer is processed using a manually defined set of rules to form UNL expressions.The root word of each important word in a sentence forms a UW and the suffixes attached to these words determine the relations between Uws. Some words including articles, prepositions, and adjectives from attributes of these UWs.Also, the UWs that have to be processed together form a hyper node or a subgraph.The hypernodes are identified using suffixes of the Tamil words and are represented by a subtree in UNL Graph.

B. UNL DeConverter

The process of converting UNL Graph to target language text is called Deconversion. A new algorithm is developed for deconversion.This algorithm traverses the UNL Graph in DFS manner and adds the UWs linearly along with few custom defined connectors. UNL attributes prefixes are added to the UWs.These UWs are modified with SimpleNLG to form tense verbs if they have attributes pertaining to the senses.Then a DFS is performed on the graph to create the sentence from it by joining the words together connected by words representing the relations.

C.Evaluation of emotional speech

A survey was conducted to determine how well emotion was ex-pressed in the final prosodically enhanced output sentences. Six record-ings comprising of outputs expressing angry, happy and sad emotions were played and the corresponded responses were recorded.

Emotion	Percentage of total responses that accurately identified the emotion from modified voice
Anger	65%
Happiness	57%
Sadness	81%

Table 6.3 Survey Results on Identifying the Emotion present in Modified Voice

Emotion	Average score for the identified emotion given by responder on a scale of 1-5
Anger	3.6
Happiness	3.2
Sadness	4.1

Table 6.4 Scores for Emotion present in the Modified Voice

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

IV. EXISTING SYSTEM

A.Matching
The corpus is searched for a sequence of words in the source text and if there is a match, its score is incremented. i.e, greater the number of matches of the source text, greater will be its score. Those word sequences that have a score lesser than some threshold value are discarded from the corpus.

B.Transfer
Once the word sequence has been identified, the corresponding sequence is extracted from the target language. The total score is to be calculated as:

Total = Translation Score Match Score

C.Recombination
All translated word sequences are concatenated and the best N possibilities of recombinations which give the entire translated sentence are chosen.Y. Choueka Bar, et al discuss the application of the example-based approach for Arabic to English translation[9]. But this approach has few drawbacks as it gives way for many exceptions to arise, for which new rules need to be formulated.Also, according to Yves Lepage, et al., pure example based approach is feasible only by adopting proportional analogy.

V.NIST

The NIST 2005 Machine Translation Evaluation (MT-05) was part of an ongoing series of evaluations of human language translation technology.BLEU measures translation accuracy according to the N-grams or sequence of N-words that it shares with one or more high-quality reference translations. But it might be better to weigh more heavily those N-grams that are more informative – i.e., to weigh more heavily those N-grams that occur less frequently, according to their information value. This helps to combat possible gaming of the scoring algorithm since those N-grams that are most likely to (co-)occur would add less to the score than less likely N-grams[24].

In f o (w1:::wn) = log2 the # occurrences o f w1:::wn 1 (6.4)

n=1 >å (1) > { [Lre f] }

N 8 å In f o (w1 : : : wn) 9 exp Lsys

Score = å all w1 :::wn that co occur b log2 min ; 1

> 1n > : occur in sys out put ;

< all w :::w that co (6.5)

where

b is chosen to make the brevity penalty factor = 0.5 when the no of words in the system output is 23 rds of the average no of words in the reference translation

N = 5

Lre f = the average number of words in a reference translation averaged over all reference translations

Lsys = the number of words in the translation being scored.

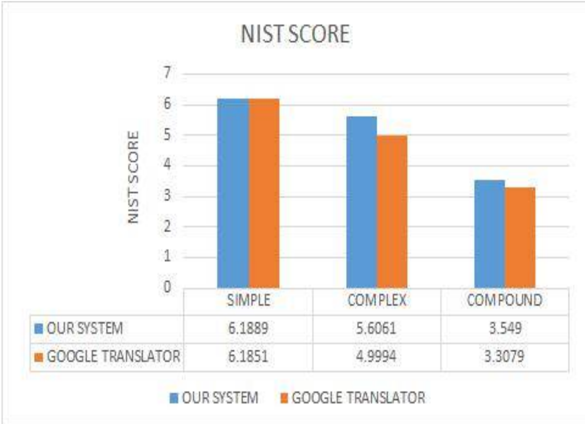
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017



The NIST score comparison of our System to that of the Google translator is shown in Figure 6.2.

VI. METEOR

METEOR stands for Metric for Evaluation of Translation with Explicit Ordering.It is based on the harmonic mean of unigram precision and recall, with recall weighted higher than precision. It also has several features that are not found in other metrics, such as stemming and synonym matching, along with the standard exact word matching. The metric was designed to fix some of the problems found in the more popular BLEU metric, and also produce the good correlation with human judgment at the sentence or segment level.It involves alignment of the reference text with the translated text.The score is computed as

M = Fmean(1   p) (6.6)

where M denotes the score and p denotes the penalty computer as

p = 0:5 C (6.7)

um3

where c is a number of chunks of data, um is the number of unigrams that have been mapped.The METEOR score comparison of our System to that of the Google translator is shown in Figure 6.3. Fluency is a measure of the grammatical correctness of the sentence.Adequacy, on the other hand, determines the degree of information conveyed by the translated text.Both measures are to be evaluated manually through a survey. The text translated by our system along with Google translation of the same text and the source text is given to a group of people who knew both the languages.The responses are shown in Figures 6.4 and 6.5.

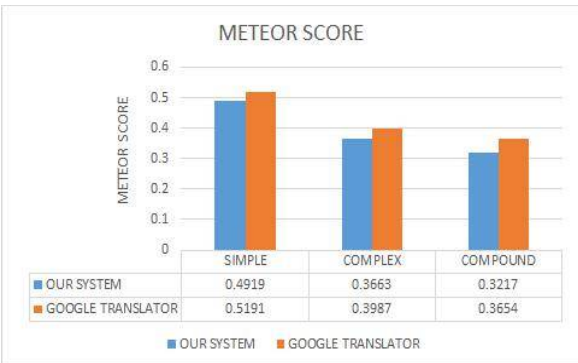


Figure 6.3 METEOR scores: Our system vs Google translator

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

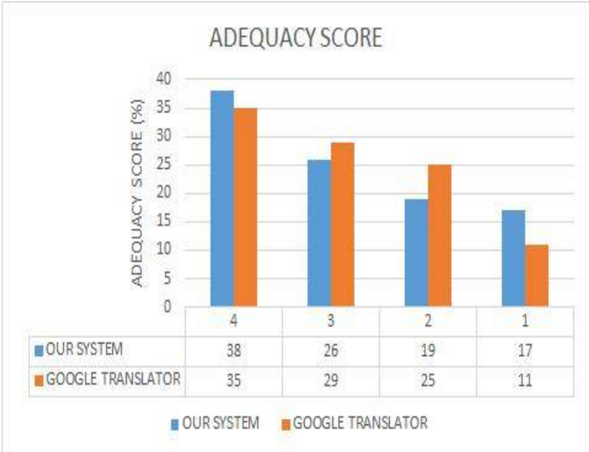


Figure 6.5 Adequacy scores: Our system vs Google translator

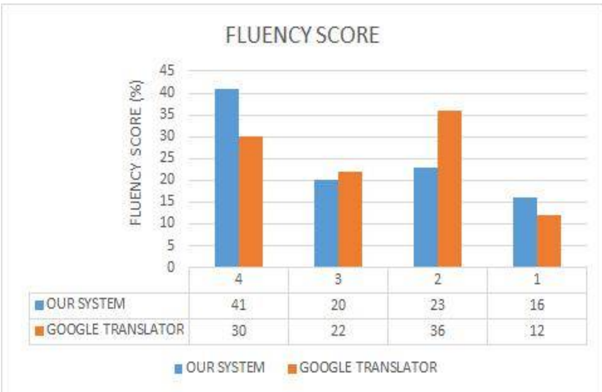


Figure 6.4 Fluency scores: Our system vs Google translator

VII. FUTURE WORK

By increasing the word count in our Wordnet and also adding more domain information to the existing words, better accuracy could be achieved.Also, implementing Named Entity Recognition would aid in better POS Tagging.Few more POS Tags have to be added and corresponding rules need to be changed. Wordnet has to be improved with more words and word domain information needs to be incorporated for each word.Based on the prototype for these three emotions, a lot more emotions can be handled. This system is scalable with these improvements and can play a vital role in the field of Natural Language Processing.

VIII.CONCLUSION

This system translates Tamil text to Tamil voice with emotions .In the case of a morphologically rich language such as Tamil, the use of an interlingua is better for translation.A custom built morphologically analyzer is used to extract noun, intense and gender from the Tamil text. A POS Tagger was also created to classify the winter noun, verb, object, adverb, and adjective. Based on the POS tags and morphological structures rules are devised to convert the Tamil text to UNL with an enconverter.A Decon-inverter, with a newly developed algorithm, builds a simple Tamil voice from the UNL text with the help of Simple NLG.Then the emotion in the sentence is identified from the UNL and based on the identified emotion, the TTS output for the translated English text is modified to incorporate prosody on it. The results

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

of the evaluation showed that UNL is indeed efficient in translation.We have successfully added three emotions (sadness, happiness, and anger) to the machine voice.

REFERENCES

[1] G. G. Chowdhury, "Natural language processing," *Annual review of information science and technology*, vol. 37, no. 1, pp. 51–89, 2003.

[2] K. W. Church and R. L. Mercer, "Introduction to the special issue on computational linguistics using large corpora," *Computational linguistics*, vol. 19, no. 1, pp. 1–24, 1993.

[3] J. Lehrberger and L. Bourbeau, *Machine Translation: Linguistic characteristics of MT systems and general methodology of evalua-tion*, vol. 15. John Benjamins Publishing, 1988.

[4] L. Dugast, J. Senellart, and P. Koehn, "Statistical post-editing on systran's rule-based translation system," in *Proceedings of the Sec-ond Workshop on Statistical Machine Translation*, pp. 220–223, 2007.

[5] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based trans-lation," in *Proceedings of the 2003 Conference of the North Amer-ican Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pp. 48–54, 2003.

[6] V. Kumar, R. Sridhar, S. Narayanan, and S. Bangalore, "Enriching spoken language translation with dialog acts," in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguis-tics on Human Language Technologies: Short Papers*, pp. 225– 228, 2008.

[7] D. Chiang, "A hierarchical phrase-based model for statistical ma-chine translation," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 263–270, 2005.

[8] A. Ramanathan, J. Hegde, R. M. Shah, P. Bhattacharyya, and M. Sasikumar, "Simple syntactic and morphological processing can help english-hindi statistical machine translation.," in *IJCNLP*, pp. 513–520, 2008.

[9] K. Bar, Y. Choueka, and N. Dershowitz, *An Arabic to English example-based translation system*. PhD thesis, 2007.

[10] Y. Lepage and E. Denoual, "Purest ever example-based ma-chine translation: Detailed presentation and assessment," *Machine Translation*, vol. 19, no. 3-4, pp. 251–282, 2005.

[11] J. G. Carbonell, S. Klein, D. Miller, M. Steinbaum, T. Grassiany, and J. Frey, "Context-based machine translation," pp. 19–28, 2006.

[12] H. Uchida, M. Zhu, and T. Della Senta, *Universal Networking Lan-guage*. 2005.

[13] P. Kumar, *UNL Based Machine Translation System for Punjabi Language*. PhD thesis, Tharapur University, Patiala, 2012.

[14] D. Schwarz, "Data-driven concatenative sound synthesis," in *Pro-ceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy,2009*.

[15] R. Carlson, T. Sigvardson, and A. Sjölander, "Data-driven formant synthesis," *KTH, Stockholm, Sweden, Progress Report*, vol. 44, pp. 121–124, 2002.

[16] J. Stark, B. Lindblom, and J. Sundberg, "Apex an articulatory syn-thesis model for experimental and computational studies of speech production," *TMH-QPSR*, vol. 2, pp. 45–48, 1996.

[17] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Ki-tamura, "Speech parameter generation algorithms for hmm-based speech synthesis," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 1315–1318, 2000.

[18] R. J. McAulay and T. F. Quatieri Jr, "Computationally efficient sine wave synthesis for acoustic waveform processing," June 26 1990. US Patent 4,937,873.

[19] A. W. Black and K. A. Lenzo, "Flite: a small fast run-time synthesis engine," in *4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis*, 2001.

[20] S. Sangeetha and S. Jothilakshmi, "Syllable based text to speech synthesis system using auto associative neural network prosody prediction," *International Journal of Speech Technology*, vol. 17, no. 2, pp. 91–98, 2014.