Automatic Tamil Lyric Generation Based on Image Sequence and Derived Tune

Rajeswari Sridhar
Department of Computer Science and Engineering
Anna University
Chennai 600025
rajisridhar@annauniv.edu

Janani Damodaran
Department of Computer Science and Engineering
Anna University
Chennai 600025
jananidamodaran@gmail.com

Nivaedha Dharmaraj
Department of Computer Science and Engineering
Anna University
Chennai 600025
nivae.dharmaraj@gmail.com

Srividhya Sampath
Department of Computer Science and Engineering
Anna University
Chennai 600025
srividhya.sampath@yahoo.com

Abstract- **The developed system is intended to generate Tamil lyrics from sequence of images, and derived tune. The choice of tune is automatically identified from the input situation and the notes are generated using a newly devised algorithm based on Carnatic music characteristics. The system, thus, generates two set of lyrics, based on the input: 'Context (from Text) and Tune' and 'Context (from text) and Image'. Image sequence helps in generating lyrics that are in accordance with the visual effects of the song. This is achieved by object extraction from the input images and location identification using a heuristic algorithm. Singable lyrics are generated by creating tune using Carnatic raga lakshanas which are its characteristics. An appropriate raga for the extracted emotion is determined by referring to a designed Raga-Emotion database. The raga characteristics are then used to synthesize notes leading to tune.**

Index terms: **Natural Language Processing, Lyric Generation, Image Identification, Raga synthesis**

I. INTRODUCTION

Automating the process of lyric generation can serve many purposes, from being an aid for lyricist to generating lyrics in the absence of lyricist. Automatic or Manual, Lyric generation can be done based on (a) situation and (b) based on tune [1]. In order to generate lyrics automatically, both the methods have their own drawbacks and advantages. In this work, we are combining the two methods and adding location for automating the process of Tamil lyric generation.

There are many challenges in automating the process of lyric generation [2] as the lyrics need to be meaningful and semantically related to a scenario. Moreover, there should be intra-line and inter-line logical coherence of words. The choice of words contributes to the quality of lyrics. If one looks at the evolution of the word composition of the Tamil film lyrics, it becomes quite evident that off late there is a mixture of non-Tamil languages (especially English) forming a decent percentage of word composition, which caters to the younger generation. Thus in order to be a popular and interesting lyric, the output of an automatic lyric generation system need to cater to interest of all age groups. This paper is organised as follows:

Section 2 discusses the previous work on lyric generation. Section 3 deals with the system overview and the various modules in the system architecture. Section 4 focuses on the results. Section 5 gives the conclusion and section 6 explains the future work, and extensions to the system.

II. BACKGROUND

There are a few research papers which serve as the base idea of the project. A corpus-based poetry generation system which uses templates to construct poems according to given constraints on rhyme, meter, stress, sentiment, word frequency and word similarity has been implemented [3].

One of the works for generating lyrics for Tamil [4], focuses on creating meaningful lyrics given a melody in ABC which is the 'asai' pattern specific to Tamil language. This is based on the grammar representation of Tamil language namely KNM which indicates the syllables as 'Kuril', 'Nedil', 'Mei'. The melody is analysed and a series of possible syllable patterns is generated in KNM representation scheme. The paper [5] conveys the idea of creating new poems by mixing existing poem lines based on the topics and content of the input message. It includes 3 steps viz., Word Sense Disambiguation, Topic Summarization, Genetic algorithm.

Another work on Tamil lyric generation proposes an N-gram based methodology [2]. It includes the ontological interpretation of the scenario, and the selection of the appropriate tri grams for generating the lyrics.

For generating lyrics from images we referred to the work where [6], the authors propose an automatic music image generation that extracts features from lyrics such as an event and a place using the probabilistic parametric mixture model and the scene composition knowledge. It deploys a

978-1-4799-8047-5/15/$31.00 ©2015 IEEE

technique that automatically generates a music image that matches the content of lyrics.

Another work [1] addresses the challenging task of automatically composing lyrical songs with matching musical and lyrical features. It considers one of the two following sequential approaches viz.,(a) First write the lyrics followed by music composition (b) First compose the music and write matching lyrics.

In another work [7], the authors identify emotion from a given scenario and use this emotion as a seed word to interpret the context of the scenario. A lyric model based on tri-gram is constructed which is referred using the identified seed word to generate lyrics. The lyric model, tri-gram of words, Tamil sentence rules and suffixes are used by a Morphological generator to generate lyrics.

Based on the advantages and drawbacks in the existing system, we have adopted some work and made the following contributions.

- The work described in [3] is referred, for it uses poetic parameters to generate a corpus based poetry generation to reflect the mood.
- The work of Rajeswari Sridhar et al [7] [2] did not consider tune for generating lyrics. Since, lyrics are poetry that can be sung, we modified the work of A.Anantha Ramakrishnan and Sobha Lalitha [4] to generate the lyric from the tune. However, in our case, we generated the tune in "SRGMPDN" – Carnatic music representation and later converted to Asai format.
- The work in [6] generates images from lyrics. We referred this paper for our proposed system to generate lyrics from images (Reverse idea)

III. SYSTEM OVERVIEW

The overall architecture that highlights the new contributions proposed is given in Fig. 1. The system focuses on automating the lyric generation process, given an input scene description and image in which the scene is proposed to be shot. The system flow begins at morphological analyzer, which stems the words and provides the suffixes and Part of Speech tags. The mood and domain are extracted from the input scene. Then, the sentence-wise emotion is recorded by making use of emotion-bearing keywords. The mood is used to select the appropriate raga. The grammar of the raga is then used to generate the notes based on the Ragalakshana. The notes appropriately capture the "rasa" of the raga. Also, causal effect relations are extracted based on observed patterns in sentences, aiming to understand the epitome of the input scene description [8]. This results in seed words, which contribute to the words in the final lyric. The emotion and domain pertinent to the scene play a vital role in choice of words in the lyric.

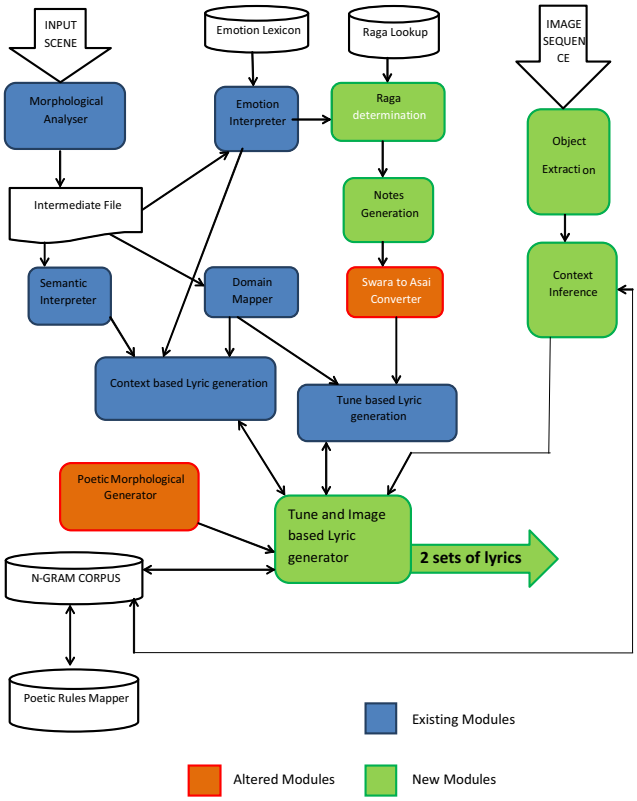

Fig. 1: Overall Architecture

A. *Raga Determination*

Emotions determined in the emotion interpreter module are given as the input to the raga identification module. To identify the emotion, the work of [7] is used. The list of ragas for these emotions is identified from Carnatic music literature and is stored in our designed database. This database compared and the raga is retrieved that matches a given emotion. We find the raga that is common for all the emotions. If there is no common raga, we propose the use of the Longest Common Subsequence (LCS) algorithm [9]. In the event where multiple ragas cater to a single emotion, the LCS of the Aarohana of these ragas are identified to determine the raga for an emotion.

B. *Notes Generation*

Each raga has several 'Raga Lakshanas' which shows them uniquely. The system considers the Raga Lakshanas like GrahaSwara (beginning note of a Raga) and Amsa (Characteristic Phrases). The generated notes start with the grahaswara, which is the starting note of the raga. Then subsequent notes are generated based on the previous note. A range of four next possible notes are considered for each swara from the raga's aarohana and avarohana. Characteristic Phrases, which adds beauty to the raga, and that distinguishes one Raga from the other is considered for generating notes. If more than one characteristic phrase is available for a previous note, then any one is randomly picked. The system selects from the next possible swara and the characteristic phrases

randomly and inserts into the notes. In this work, we have handled Roopaka Thalam and hence the generated notes correspond to the 'Roopaka' thalam. So each line has atmost 12 notes corresponding to 3 counts of Roopaka Thalam. Thus for 8 lines, atmost 12 *8=96, notes are generated.

C. Object Extraction

Typically lyrics – a poem that is singable, are typically used for a song in a movie and hence will have a background location. In this work, we have generated lyrics to incorporate the location of the song by including words of the background in the first few lines of the lyric. This module is used to extract the objects from an input image. Given an image or a sequence of images, various objects in each of them are identified by using RGB color values. In this work, the objects considered are sand, sky, rock, water, leaves, bark and snow. Conflicts are avoided by heuristics results obtained from trial runs.

D. Context Inference

In order to identify the context of song's shooting location, context of the scene need to be identified. This context will also contribute in choosing words for generating lyrics. The list of objects extracted in the object extraction module is read from a file. For each category of context stored in the database, we retrieve the objects and determine the count of extracted objects that match with the retrieved objects. The category that has the maximum match is identified as the context of the sequence of images.

E. Lyric Assembler and Generator

Using the above designed algorithm and using the existing algorithm, we generate four types of lyrics as follows:

Context based lyric[7]: We add emotion, domain, dialect and seed words to the set of rules and generate lyrics by considering only the scene.

Context and Tune based lyric: We design the tune using our algorithm and use existing algorithm to generate lyrics from tune [1] by adding emotion and domain.

Context and Image based lyric: We add location, emotion, domain and asai to the set of rules and generate lyrics. This module generates a lyric by utilizing poetic line generator [10]. The poem is ensured to possess unique lines and different poems are generated each time the same scene given as input. Moreover, this module provides the appropriate word selection rules, based on the user selection, i.e. scene-based, image-based, tune-based or pa-based. Further, to make the lyric concur with the input scene, lines are generated that begin with seed words collected from the semantic interpreter [2].

IV. RESULT ANALYSIS

The lexicon designed by referring Tamil poems, existing movie lyrics, scrubbed and cleaned EMILLE corpus, Tamil books like Kamba Ramayanam, Nalayira Divya Prabhandham, Bharathiyar Kavithaigal and English synsets comprises of roughly 13,000 unique words. 40 different ragas that reflect emotions namely love, happy, sad and courage are used in this work. The developed lexicon is categorized according to domains mentioned above. Corpuses are built separately for different types of lyrics in each domain. Quadragram corpus for scene, pa and tri-gram corpus for tune and image are created from the lexicon. Around 200 scenes taken from Tamil short stories and 50 images referring to forests, snow, sea and desert are used for testing purpose. 140 images are used for training. The results of the training are then used to develop the algorithm.

Input scene is,

அவன் ஒரு போர்வீரன். அவன் அவளை காதலிக்கிறான். அதனை அவளிடம் சொல்கிறான். அவளின் பதிலை எதிர்பார்த்து காத்திருக்கிறான். அவள் அவன் வீட்டிற்கு ஒரு திருமணத்திற்காக செல்கிறாள். அங்கு அவள் தன் காதலினை ஆடியும் பாடியும் சம்மதிக்கிறாள். உடனே அவன் மகிழ்ச்சி அடைகிறான்.

The input scene is processed and the POS tag for root words is obtained, from which emotion and the domain are determined. For the above specified input scene, Emotion and Domain are found to be **'Happy'** and **'Love'** respectively. The determined emotion and domain aids in generating the scene based lyric as,

நால்விழி சேர்ந்து நாடகமாடியது காதலில்
இமை பிரிந்தது உறக்கம் இன்றி
சுகம் தான் தோன்றுதே பார்வையில்
நினைவுகளின் ஒளி ஆனாய் மின்னலே

Naalvizhi sernthu naadagamaadiyadhu kaadhalil
Imai pirindhadhu urakkam indri
Sugam dhan thondrudhe paarvaiyil
Ninaivugalin oli aanai minnale

The appropriate raga for 'Happy' and 'Love' is determined as **'Saranga'** by looking up into the 'Raga Lookup Database'. The notes generated for the identified raga is,

s–P MpDN SD-P
Ms-P Mp-D NSD-
s–P MpDN SD-P
Ms-P Mp-D NSD-
P-d- N-dn S-p-
N-NS -MmD -nnd
-S-N pM-m -s-P
MP-P -DNS D-PM

The generated notes are transformed to the KNM format. The asai pattern obtained from the KNM format is then used to pick appropriate words from the tri-gram corpus for tune and form lyrics as,

Oru parvai podhum
Nenju kulle ponaai

Badhil enna kannae
Anbil kolai ponen
P-d- N-dn S-p-
uyir kadhal nee
oli veesum mugham
ullam adhil mazhai
For image based lyrics, the sequence of input images are given Fig 2,

Fig. 2. Sequence of Input Images

The objects from the input images are extracted and the location is inferred as **'Desert'**. The lyrics pertaining to the location is obtained as,

உன் கண்ணசைவில் பாலைவனம் பூப்பூக்கும்
கூலாகும் இப்பூமி உன் கால்பட்டால்
Unkannasaivil paalaivanam poopookum
Coolaagum ipboomi un kaalpattal
The performance of the system is measured using the parameters like,

A. Assortment of Words

It scores a poem based on the number of unique words in the poem [7]. Hence, ratio of the number of distinct words to the total number of words in the lyric is taken as assortment of words.

$$\text{Assortment of words} = Dw/Tw * 100$$

where, Dw= number of distinct words in lyric
Tw = total number of words

Scene based lyrics has high assortment of words value, because, there are no restrictions except for the number of words in a line. Hence, wide set of words are available. Since Pa based lyrics are obtained by re-engineering scene based lyrics, they also have high assortment of words value. There is only limited number of words that characterises the features of a location. Thus, image based lyrics has less assortment of words value when compared to scene and pa based lyrics. The generated notes are converted to asai pattern and words are picked accordingly. Hence, there are only less number of words adhering to the asai pattern, and so, tune based lyrics have the least assortment of words value.

B. Semantic Relatedness

This measures the extent to which the engendered lyric concurs with the input scene. This is measured by obtaining feedback from end users, who provide a score to the lyric.
Scene Based Lyrics are generated purely based on the emotion and domain of the input scene, so they are highly related to the input scene. Since Pa Based Lyrics are obtained by re-engineering Scene Based Lyrics, they also concur highly with the input scene. Constraint on the asai pattern reduces the semantic relatedness of the Tune Based Lyrics. Only two lines of Image Based Lyrics are generated and to reflect the crux of the input scene in a limited manner than Scene and Pa Based Lyrics.

C. Musical Score

Poetic components like edhugai, monai, iyaibu are taken into account to assign a musical score to a poem automatically [11]. For the tune-based lyric, the musical score is high, as it is ensured to be singable in the input tune.
For tune based lyric,

$$\text{Musical Score} = 50 + (RC/TW) * 100$$

For other lyrics,

$$\text{Musical Score} = (RC/TW) * 100$$

where, RC - Number of rhythmic components (edhugai, monai, iyaibu) and
TW - Total number of rhythmic components in lyric.

We are adding more weight to tune based lyrics since lyric is generated based on the asai derived from the notes generated which enhances the musicality of the final lyric.
Since only the Tune Based Lyrics are generated based on the asai format of the generated notes, they have high musical score than other lyrics.

D. Logical Coherence

Coherence is a measure of flow of words in a piece of writing. Coherence between words in a line is ensured based on the n-gram corpus, framed by mining various poems. This is measured via survey from users of age group 18-25.
Constraints on the asai pattern reduce the coherence between words in Tune Based Lyrics. And hence they have less logical coherence value than other lyrics.

The results of these parameters are shown in Table 1

TABLE 1. RESULTS OF PARAMETERS

Lyric	Assortment of Words	Semantic Relatedness	Musical Score	Logical Coherence
Scene Based	98%	98%	80%	98%
Tune Based	90%	96%	95%	93%
Image Based	94%	95%	85%	95%

At the end of each stage the results were analyzed to ensure better performance of the system. The following describes the evaluation methodologies dealt for each one of them,

E. Notes Generation

The survey is conducted to rate the notes generated for a raga determined by the system. A survey is conducted where a music teacher was made to evaluate the notes generated. From Fig.2 and Fig. 3, we can observe that the

survey shows 80% of the notes generated are in conformance with the raga and 83% are singable.

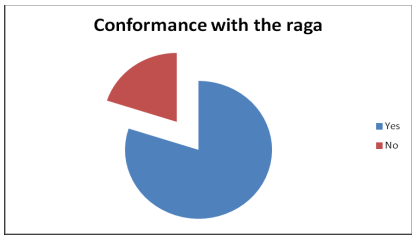

Fig. 3. Conformance with the raga

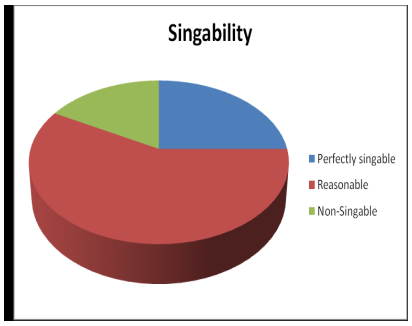

Fig. 4. Singability

The grammar of the raga involves incorporation of several features. Since the system has handled only few, there is a reduction in percentage conformance of the notes generated to the respective raga rules. Also the system generated notes in few cases cannot be sung as-is. This is attributed to the fact that the octave scales for each swara is randomly picked.

F. Image Identification

Manual Evaluation of whether the image identified by the system is correct. Identification accuracy is a parameter which is used to measure how accurately the image is identified by the developed algorithm.

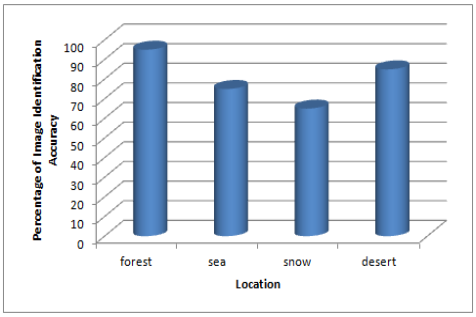

Fig. 5. Image Identification

The system is designed to identify bark, leaves, sky, water and sand. These objects are then mapped to the most appropriate location. The training and testing is done using 190 images containing the aforesaid objects, using which the heuristics are developed. Texture based differentiation and identification is not handled. And that's the reason for the reduction in percentage of identification accuracy of the locations considered

G. Location Based Lyric's Adherence to image

Feedback from end user is obtained. The audience were asked to match the location identified by the system and the lyric generated based on that location. From Fig. 6 and Fig.7 we can observe that the feedback reveals 95% of the generated lyrics are relevant to the location and creates visual imagery of the location.

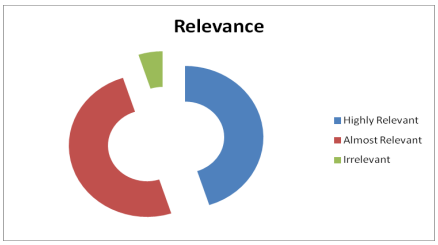

Fig. 6. Relevance

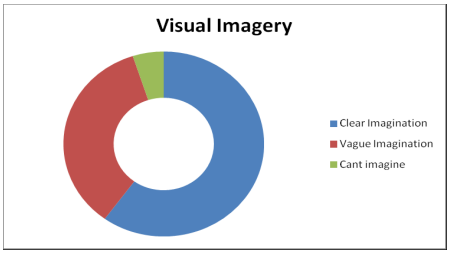

Fig. 7. Visual Imagery

Corpus is built for each domain namely love and friendship (optimistic, pessimistic), nature, patriotism. Not all objects present in the location identified are referred in the generated lyric. It is because, there is a constraint on the number of lines generated that refer to the location (only 2).

H. Tune Based Lyrics

It is evaluated by comparing with existing tune based lyric generation. It measures the pleasantness, singability, tunes novelty and mood reflection.

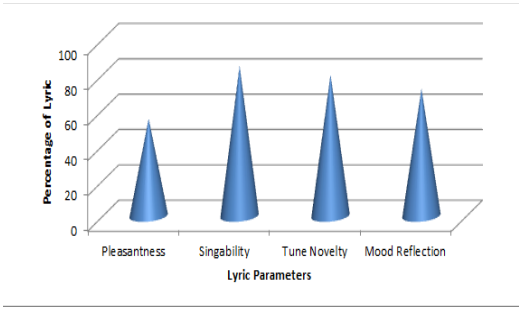

Fig. 8. Tune Based Lyrics

The system auto-suggests the tune in which the lyric can be sung. The percentage of mood reflected is comparatively lower than other parameters viz., singability and novelty, because the feel of the raga will be effective only when one hears it in an unperturbed state of mind, which isn't always possible. Pleasantness is the least of the three since there is a stringent restriction as to the use of the words. The swara to asai conversion imposes a constraint where only 6 asai patterns are allowed, which obviously limits the choice of words. Singability is assured to a decent extent, for the lyrics are generated from the asai, which is derived from the raga notes.

V. CONCLUSION AND FUTURE WORK

The current system generates lyrics based on input situation (text), images and tune. The system does not add any rhythmic components to the lyric. Also, emotion, which is identified from input, is based on keywords. Therefore, poem will not be generated if there is no emotion bearing keyword in the input scene. Images with more pixel values not corresponding to the correct distinguishing feature will give wrong location as result, so the final lyrics produced will not be significant. Only RGB color values are considered for object identification. At most 2 locations are handled for a scene. Determining raga for more than 2 emotions for an input scene is not handled. Other emotions like patriotism, peace, etc., are not considered. For determining a raga for 2 emotions, Longest Common Subsequence (LCS) of aarohana was only considered. Other features of a raga are not considered. Other thalams like adhithalam etc are not handled. Four notes for a word is fixed.

The present system identifies direct emotion keywords. It can be extended to identify indirect emotions from the scene by making the system to comprehend the situation. Further, the system produces incorrect lyrics if the scene contains words from multiple domains. Thalam can be made a dynamic/ user specified feature. This also involves variation in tempo which has to be kept in mind when the lyric is sung. Additional nuances (Raga lakshanas) peculiar to a raga can also be incorporated while generating notes, so that it reflects the raga rasa. RGB combined with texture based scene understanding can result in better identification of the shooting location. Consequently, the lyrics obtained will be still more relevant.

REFERENCES

[1] Jukka M. Toivanen, Hannu Toivanen, and Alessandro Valitutti, "Automatical composition of lyrical songs", In Proceedings of the Fourth International Conference on Computational Creativity, Sydney, Australia, pp. 87–91, 2013.

[2] Rajeswari Sridhar, D.Jalin Gladis, Ganga Kameswaran, and G.Dhivya Prabha, "Automatic tamil lyric generation based on ontological interpretation for semantics", In Sadhana - Academy Proceedings in Engineering Science, Vol. 39, No. 1, pp 97 - 121, 2014

[3] Simon Colton, Jacob Goodwin, and Tony Veale, "Full face poetry generation", In proceedings of the Third International Conference on Computational Creativity, pp. 95–102. ACM, 2012.

[4] A.Anantha Ramakrishnan and Sobha Lalitha, "An alternate approach towards meaningful lyric generation in tamil", In Proceedings of the NAACL HLT 2010 Second Workshop on Computational Approaches to Linguistic Creativity, pp. 31–39, 2010.

[5] Owen Noel Newton Fernando, Adrian David Cheok, Nimesha Ranasinghe, Kening Zhu, Chamari Edirisinghe, and Yan Yan Cao, "Poetry mix-up: A poetry generating system for cultural communication", In Proceedings of the ACE 2009 International Conference on Advances in Computer Entertainment Technology, pp. 396–399, 2009.

[6] Go Kikuchi and Hiroyuki Kasai, "Lyrics-based automatic music image generation using scene knowledge for music browsing", In Proceedings of the 2012 IEEE International Conference on Consumer Electronics(ICCE), pp. 249–250, 2012.

[7] Rajeswari Sridhar, D.Jalin Gladis, K. Ganga, and G.Dhivya Prabha, "N-gram based approach to automatic tamil lyric generation by identifying emotion", In Proceedings of ICAdC, AISC 174, pp. 919–926, 2012.

[8] Subasree Venkatsubhramaniyen, Suba Rashmi S and Rajeswari Sridhar, "Automatic Singable Tamil Lyric Generation for a Situation and Tune by Interpreting Scenario Based on Causal Effect", Under Revision I at Sadhana Journal.

[9] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, "Introduction to Algorithms (3. ed.)", MIT Press ;2009

[10] K.V. Madhavan, S. Nagarajan, and Rajeswari Sridhar, "Rule based classification of tamil poems", In International Journal of Infor- mation and Education Technology, pp. 156–158, 2012.

[11] Subasree Venkatsubhramaniyen, Subha Rashmi, and Rajeswari Sridhar, "Classification of tamil poetry using context free grammar using tamil grammar rules", In Third International Conference on Computer Science Information Technology (CCSIT 2013),Computer Science Conference Proceedings (CSCP), pp. 159–167, 2013.

[12] Simone Hill, "Markov melody generator", In Fall 2011 at University of Massachusetts Lowell, 2011.

[13] Rajeswari Sridhar, R Rajkiran, N J Narendra Kumar, and M Giridhar, "Recognition of kalippa class of tamil poetry", In 12th Inter- national Tamil Internet Conference, pp. 22–27, 2013.