Computer Interpreter for Translating Written Sinhala to Sinhala Sign Language

M. Punchimudiyanse*[1] & R. G. N. Meegama[2]

[1]*Department of Computer Science, Faculty of Natural Sciences, The Open University of Sri Lanka*

[2]*Department of Computer Science, Faculty of Applied Sciences, University of Sri Jayewardenepura*

Abstract

Sinhala Sign Language (SSL) is the preferred medium of communication amongst the deaf community of Sinhala. A conversational sign in SSL represents a word or phrase in Sinhala whereas a fingerspelling sign represents a character in the Sinhala alphabet. A word in Sinhala has different word forms based on tense and case structure of the Sinhala language. In the early stages of SSL evolution, a single SSL sign represented a base Sinhala word and all its word forms. This technique has deprived essential word form information to a deaf person. The SSL research community has come up with a technique of showing few characters of the expected word form in fingerspelling as prefix and/or suffixes with the sign of the base word to solve the problem. This research paper presents a database driven translator, which implements the said technique by translating written Sinhala sentences into SSL sequences. It gives the opportunity to the SSL research community to popularize this translation technique among SSL users. Moreover, a 3D avatar built earlier for animating spoken Sinhala in SSL is used to animate the translated SSL sequence. A test sample of 100 sentences and a word form database of 500+ entries to match 50 conversational SSL signs (30 nouns and 20 verbs) are formed to test the translator. The database driven search mechanism provided 100% successful translator performance while having 70.83% average sentence identification rate for six SSL users including three professional SSL interpreters.

---

* Corresponding Author: M. Punchimudiyanse; email: malinda@ou.ac.lk

ORCiD: http://orcid.org/0000-0001-5790-0761

Introduction

Sinhala and Tamil languages are the two official languages of Sri Lanka. However, the aurally handicapped Sinhala community uses Sinhala Sign Language (SSL) as their preferred medium of communication. Several deaf schools scattered throughout Sri Lanka use SSL to teach the regular Sinhala curriculum to schoolchildren. The Sinhala community lacks the knowledge of SSL, thereby the services of a sign interpreter is required to communicate with a Sinhala deaf person. Since there are only a handful of professional SSL interpreters available in Sri Lanka, deaf people face various obstacles in their day-today interactions with the Sinhala speaking community.

Traditional SSL does not have sufficient vocabulary of grammar rules to accommodate complex grammar of written Sinhala. The written Sinhala has 9 cases, 3 tenses with masculine, feminine and question forms. In contrast, SSL may have a single sign gesture to represent all such word forms of a particular Sinhala noun or verb except a few exceptions for past tense verbs and the feminine form of nouns. Therefore, the computerized Sinhala to SSL translators (Liyanapathirana, 2014; Kulaveerasingam, Wellage, Samarawickrama, Perera & Yasas, 2014; Punchimudiyanse & Meegama, 2015a) are replacing all different word forms of a Sinhala word with the common sign of SSL in the same way it is done by the human SSL interpreters. However, this translation technique does not express the exact meaning of written Sinhala sentences to a deaf person due to missing cases and tenses, and it is a serious problem which contributes to the low pass rates among deaf schoolchildren in school examinations.

The SSL community has a different SSL signing technique to solve the written Sinhala to SSL translation problem. This technique is used in supportive textbooks of SSL (Siyambalagoda et al., 2004) that match the regular school curriculum. The said technique uses a combination of a prefix, sign for base word and a suffix to visualize exact Sinhala word form in SSL. For example, to sign the Sinhala word තාත්තාගේ (meaning : father's), the SSL uses the sign for the base word තාත්තා (father) and follow it up with two fingerspelling signs "ග්, ඒ" to show the deaf person that it is තාත්තා +ගේ (තාත්තාගේ).

This research paper proposes a database driven written Sinhala to SSL translator, which emulates the said technique along with traditional Sinhala to SSL translation terminology. This translator system is tested using a 3D avatar system

(Punchimudiyanse & Meegama, 2015a) developed for spoken Sinhala to SSL translation.

Technical and Linguistic Background

Sinhala Sign Language

The SSL gestures are classified as conversational and fingerspelling gestures. Conversational gestures are preferred by the SSL community because they represents a word or an entire phrase as shown in figure 1(a) and figure 1 (b). Character by character representation (fingerspelling) is used to show the Sinhala words that are not having SSL signs such as person's name. The fingerspelling alphabet of SSL (Dias & Dias, 1996; Stone, 2007) contains 60 finger spelling signs representing the 60 characters of the modern Sinhala alphabet (NIE SL, 1989) and a fingerspelling sign for the ඦ character in Unicode Sinhala standard (Unicode.org, 2016).
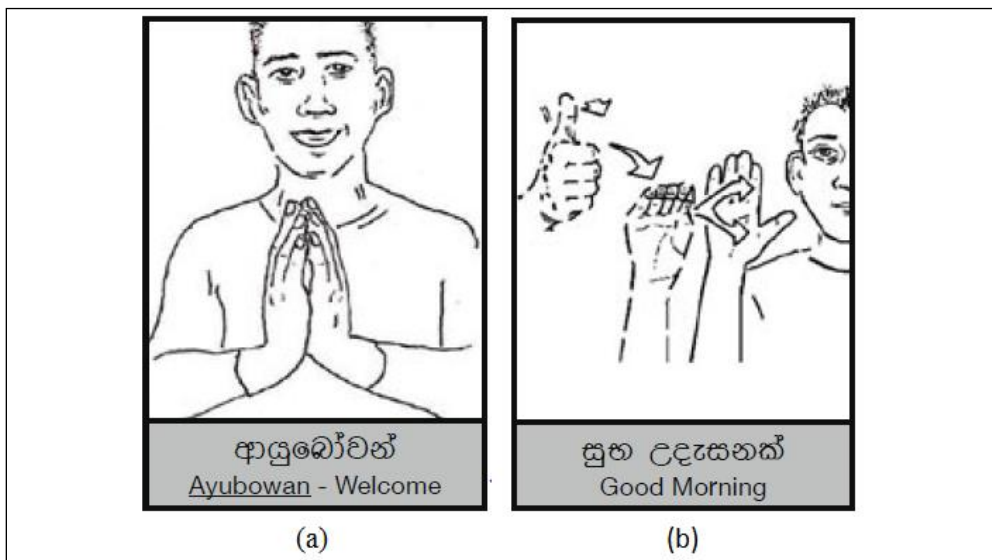

ආයුබෝවන්
Ayubowan - Welcome

සුභ උදෑසනක්
Good Morning

(a)

(b)

Figure 1: SSL sign for (a) single word (b) phrase

Fingerspelling in American Sign Language (ASL) and British sign language (BSL) are straightforward tasks of using individual character signs of an English word. For example, the word "JOHN" can be shown with four fingerspelling signs "J","O","H", "N" of ASL or BSL as shown in figure 2(a).

However, SSL fingerspelling uses the phonetic pronunciation of a word to construct the fingerspelling sequence. For example, the same word "JOHN" in Sinhala is written as "ජෝන්" and its phonetic pronunciation is "ජ්, ඕ, න්", which can be fingerspelled using three SSL fingerspelling signs as shown in figure 2 (b).
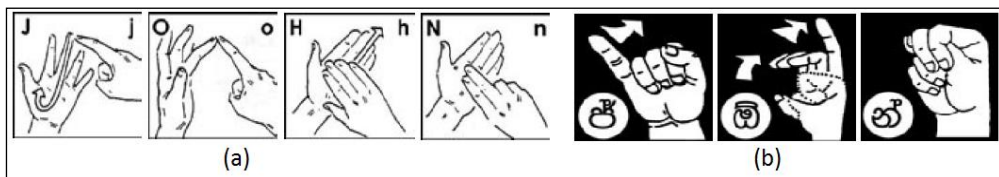
Figure 2: Fingerspelling John (ජෝන්) (a) in BSL (b) in SSL (black background)

SSL does not have sufficient vocabulary to support different word forms of Sinhala. Hence, the SSL user community has developed a strategy of signing those different word forms of Sinhala in SSL.

Techniques for Sign Language Synthesis and Animation

Sign languages differ from one another as shown in "spread the sign" web portal (European Sign Language Center, 2012). Several works on translating a spoken language into respective sign language exist for small sign vocabularies or for specialized domains. Moreover, a large number of researches on sign language synthesis were carried out for the ASL and for the sign languages of the European Union.

The popular technique employed in translation and sign synthesis is word to word mapping of signs to match words in a sentence. If a sign does not exist for a particular word, it is fingerspelled (Stone, 2007). The second technique of sign translation is using parallel corpuses where one corpus has words/phrases of a spoken language and the other corpus contains sequences of signs matching the first corpus. Natural language processing (NLP) techniques and Machine translation techniques are used in parallel corpus based systems for generating translated sign sequences from an

input sentence (Almeida, Coheur, & Candeias, 2015; De Martino et al., 2016; Efthimiou et al., 2009; Kouremenos, Fotinea, Efthimiou, & Ntalianis, 2010; Othman & Jemni, 2011;).

The word-to-word translation technique is sufficient to process most Sinhala words having a SSL sign in the traditional Sinhala to SSL translation technique, because all word forms of a Sinhala word are replaced with the base SSL sign. However, the word-to-word sequencing in sign language translation may not give the exact meaning of an input sentence (Elliott, Glauert, Kennaway, Marshall & Safar, 2008; Huenerfauth, 2008; Ong & Ranganath, 2005; Marshall & Safar, 2005). This also happens in the context of this research, because it requires introducing fingerspelling signs to the conversational SSL sequence to sign word forms of Sinhala.

Three techniques are used in sign synthesis and animation. The first technique is joining word-by-word videos of human interpreters (Solina, Krapez, Jaklic & Komac, 2000) and refinements to that technique are suggested in (Chiu, Wu, Su, & Cheng, 2007; Chuang, Wu, & Chen, 2006) for providing a smooth transition between videos.

The second technique joins motion captured animation sequences of signs to animate 3D avatars (Awad, Courty, Duarte, Le Naour & Gibet, 2009; Cox et al., 2002). A markup language SiGML (Elliott, Glauert, Kennaway, & Parsons, 2001) is used to define signs from a common set of motion captured hand postures and their transitions in the project "E-sign" (eSIGN project, 2002) that went on to translate government web pages in the British, Dutch and German sign languages. The Japanese text to Japanese sign language (JSL) translation (Kaneko, Hamaguchi, Doke & Inoue, 2010) used a TV program Making language (T2VLab, NHK Science and Technical Research Laboratories, 2011) to embed bio-vision hierarchy (BVH) based motion captured sign sequences to generate television programs that can be played in Microsoft directx based TVML player.

The third technique also uses animated 3D avatars, which shows postures of a sign in key frames and the movements between postures are mathematically calculated (Almeida, 2015; Davidson, 2006; Efthimiou et al., 2009). Huenerfauth (2006) divided a sign posture into several bone channels and animated them independently to achieve the collective movement between two postures. In this research, a 3D avatar utilizing the third technique (Punchimudiyanse & Meegama, 2015a) is used to test the translator.

Methodology

The main goal of this research is to process Unicode Sinhala sentences and translate it to SSL sequence in a mixture of traditional and written Sinhala signing techniques for word forms. For that purpose, a synonym database and a word form database are proposed. An algorithm is designed to perform the written Sinhala to SSL translation. To test the translation, an existing 3D avatar is used.

Building synonym database for the translator

A Sinhala word having an SSL sign may have a lot of Sinhala synonyms related to it appearing in an input sentence. In the traditional Sinhala to SSL translation, any word form of a synonym is replaced with the sign of the base word. For this purpose, a synonym database is built for Sinhala base words that have conversational signs defined in the sign database. The synonym database has the following entry format to denote the base word and its synonyms.

#baseword,synonym, ..,SynonymN, <carriage return>

An entry in the synonym database will look like,

#අම්මා,මව,මවුන්,මාතා,මාතෘ,මෑණියෝ,මම්මා,..,අම්මී,<carriage return>

However, the traditional translation technique is used only when a word form of synonym/base word does not have a specialized SSL signing sequence. The next section describes types of those word forms and their respective SSL sequences for the Sinhala nouns and verbs.

Building Word form database for the translator

There are nine cases (විභක්ති) in Sinhala for nouns, which can create nine word forms each for singular and plural nouns. For example, different cases that a singular noun තාත්තා (meaning: father) in Sinhala is given in Table 1. Moreover, Table 1 shows the SSL sequence for each case of the noun "father". It should be noted that the suffix in SSL sequence for a particular case is not uniform across all the nouns in Sinhala. For example, the suffix තේ is used in the noun තාත්තා in the case ආලපන to make it තාත්තේ. However, the words අයියා and මාමා have the suffixes යේ and මේ to make them අයියේ and මාමේ in the case ආලපන.

Table 1: Different cases of a 3rd person Sinhala noun (*තාත්තා* - Father)

Case	Singular	SSL sequence	English Meaning
පුථමා	තාත්තා	තාත්තා	Father
කර්ම	තාත්තාව	තාත්තා ව	Take father to
කතෘ	තාත්තා විසින්	තාත්තා විසින්	By father
කරණ	තාත්තා කරණකොට	තාත්තා කරණකොට	Because of father
සම්පුදාන	තාත්තාට	තාත්තා ට	To father
අවධි	තාත්තාගේ	තාත්තා ගේ	Father's
සම්බන්ධ	තාත්තාගෙන්	තාත්තා ගෙන්	From father
ආධාර	තාත්තා කෙරෙහි	තාත්තා කෙරෙහි	Feel about father
ආලපන	තාත්තේ	තාත්තා තේ	Call father

The situation becomes complex when a synonym with a word form that matches a noun having SSL sign exists in a sentence instead of the said noun. For example, the word form "පියාගෙන්" is a synonym word form of the noun "තාත්තා". According to the new signing technique, it has the SSL translation "පි", "තාත්තා", "ගෙන්", which is a combination of a prefix, sign for base word and a suffix. In this situation, prefix "පි" denotes "පියා", which is the synonym of "තාත්තා", and the suffix "ගෙන්" denotes the case.

When a plural noun comes for the translation, the sign for "ගොඩක්" (meaning: more than one) has to be introduced into the SSL sequence. Few examples of singular and plural synonyms with different cases are given in the Table 2.

Table 2: Word forms / SSL sequences of singular and plural synonyms

Word form in input	Base word having SSL Sign	SSL Translation	Type of noun and meaning
පියාගෙන්	තාත්තා	පි තාත්තා ගෙන්	Animate singular noun having meaning "from father"
මවට	අම්මා	ම අම්මා ට	Animate singular noun having meaning "to mother"
මෑණියන් කෙරෙහි	අම්මා	මෑ අම්මා යන් කෙරෙහි	Animate singular noun having meaning "feel about mother"
මේස වලට	මේසය	මේසය ගොඩක් වලට	Inanimate plural noun "to tables"

A verb in Sinhala is having more variations than a noun. It can have singular and plural forms, three tenses, three persons, masculine, feminine and question forms as shown in the Table 3 for the verb යනවා (meaning : go). Moreover, the Table 3 shows how a past tense verb is signed using the sign of a present tense verb followed up with the sign "finished". For example, the verb ගියා (meaning : went) becomes ගියා ඉවරයි. Furthermore, when a variation of a 1ˢᵗ 2ⁿᵈ 3ʳᵈ person is attached to a past tense verb, the last few characters of the verb should also be followed.

As it was done for a noun, if a synonym to the base verb comes in a sentence, a suitable prefix and suffix would be added to the SSL sequence to denote synonym and its word form. For example, the verb ගමන් කළෝය is a past tense plural 3ʳᵈ person synonym word form of the base word යනවා. The SSL sequence for this synonym verb could be derived as ග යනවා කළෝය or ග යනවා ඉවරයි ලෝය. These two translation variations for the word form ගමන් කළෝය indicate that it is difficult to set a single rule to determine SSL sequence multi-word word forms.

Table 3: Different cases of Sinhala verb *යයි* (meaning - go)

Type	Singular	SSL sequence	Plural	SSL sequence
Past tense				
Spoken	ගියා	යනවා ඉවරයි	ගියා	යනවා ඉවරයි
1ˢᵗ person	ගියෙමි	යනවා ඉවරයි මි	ගියෙමු	යනවා ඉවරයි මු
2ⁿᵈ person	ගියෙහි	යනවා ඉවරයි හි	ගියෙනු	යනවා ඉවරයි හු
3ʳᵈ person masculine	ගියේය	යනවා ඉවරයි යේය	ගියෝය	යනවා ඉවරයි යෝය
3ʳᵈ person feminine	ගියාය	යනවා ඉවරයි යාය	ගියෝය	යනවා ඉවරයි යෝය
Question	ගියාද	යනවා ඉවරයි ද	ගියාද	යනවා ඉවරයි ද
Non-past tense				
Spoken	යනවා	යනවා	යනවා	යනවා
1ˢᵗ person	යමි	යනවා මි	යමු	යනවා මු
2ⁿᵈ person	යහි	යනවා හි	යහු	යනවා හු
3ʳᵈ - person masculine	යයි	යනවා යී	යති	යනවා ති
3ʳᵈ person feminine	යන්නීය	යනවා නීය	යති	යනවා ති
Question	යනවද	යනවා ද	යනවද	යනවා ද

To store all such SSL sequences for different word forms of a word, the following entry format is used in preparing a scalable word-form database as a text file. Synonym, prefix segment <space> base word <space> suffix segment# <carriage return>

An entry in the word form database will look like
නිවසට,නි ගෙදර ට#<carriage return>

Design of the Written Sinhala to SSL Translator

In the traditional Sinhala to SSL translation, all word forms of a synonym or all word forms of base word are replaced with the base word of the sign. Under the written Sinhala to SSL translation technique, the respective SSL sequence of a word form of a word should replace the word. Therefore, the following algorithm is designed to process an input sentence and translate them to appropriate SSL sequence.

Algorithm

//This algorithm inputs Sinhala Unicode sentences and translates it to
//appropriate SSL sequence using both traditional and
//"base word + finger spelling" techniques

S = Input_sentence_in_Unicode_Sinhala
A = array of n-grams
SDB = synonym database
WDB = word form database

//--populate array A with all grams up to 3 grams
if three word segments exist in S then
     add each 3-gram to A
end if
if two word segments exists in S
     add each 2-gram to A
end if
add each word to A

// Search elements of A in Synonym db for a match . M stores the matched
// element. Then search M in the Word form db. WF = Word form match

For each element L in A
   Find L in SDB for a match M
  if M exists in SDB then
     B = search backwards for # from position of M to get base word B
     Search M in WDB for a match WF
    //if a WF match exist in WDB use SSL sequence of WDB. Otherwise, do
   // traditional base word replacement of the synonym
    if WF exists then
      S1 = read SSL sequence of WF from WDB
    else
      S1 = B
    end if
    //replace L in S with S1
    S = S.replace(L, S1)

    // when a phrase (gram > 1) is found, individual words in A as elements
    // should be discarded to prevent individual searches for replacements
     If word count in M > 1 then
       Replace single word elements in A matching words in M with blank
     end if
  end if
Next

// After the search and replacement process is complete S has the output //SSL sequence in Unicode Sinhala. The text that does not have //conversational signs will not be changed.

Return S

Algorithm functions as follows. An entire phrase of Sinhala can have a single SSL sign, which are searched first. For that task, the words in the input sentence are divided into 3-gram, 2-gram, 1-gram (i.e. 3-word, 2-word, 1-word) segments and searched in the synonym database from the largest segment to the smallest segment. This technique will identify synonyms (or word forms of base words), which have signs in the sign database. This search order prevents signs of individual words from replacing the parts of a phrase having a single sign.

After identifying the synonym location through the search, the base word could be obtained through a backward search for a # symbol, which denotes the base word

of the synonym. The next task is to choose whether to use the written Sinhala technique or the traditional technique of SSL translation. If the word form database has an entry to show a specific SSL sequence exists for that particular word, then that SSL sequence will replace the word/phrase of the input sentence. Otherwise, the base word replaces the word in the input sentence.

This algorithm performs the search in 3-gram to 1-gram order. Therefore, the Sinhala phrases having a single sign are translated first. There can be words within such a phrase that are having individual signs too. This may lead to partial replacement when the search goes on to replace 1-grams. To prevent such partial replacements, all the 1 grams of a translated phrase are removed from the search list.

The algorithm is implemented with Visual Basic .NET and the output is generated in the Sinhala Unicode text. However, the 3D avatar used in this research requires the input in English.

Conversion of Unicode Sinhala text to Phonetic English

The phonetic pronunciation of an unknown word is required by SSL to perform fingerspelling. The following rules are used to convert a standard Sinhala text to a phonetic sequence:

1. Introduce a hidden vowel sound "අ" after a consonant without a modifier
2. Modifiers of all consonants are replaced with the vowel sound related to respective modifier.

For the compatibility with the 3D avatar, the phonetic sequence in Sinhala is converted to phonetic English by replacing each character with a label (tag) having English letters as listed in Table 4. The "zwj" abbreviation is used for the non-printable character zero width joiner (U+200), which is used in writing special modifiers in Sinhala Unicode text such as "rakaransha", "yansaya" and "repaya".

Table 4: Tags used in Unicode Sinhala to phonetic English conversion

| Ch	Tag	ch	tag	ch	tag	ch	tag	ch	tag
අ	A	ආ	axa	ඈ	xcae	ඇ	aeae	ඉ	i
ඊ	ixi	උ	U	උෟ	uxu	එ	e	ඒ	exe
ඔ	O	ඕ	oxo	ඍa	zri	ඎaa	zrii	ඐ	zilu
ඐෟ	ziluu	ඖ	xau	ඓ	ai	ක්	k	ඬ	zk
ග්	G	ඝ්	Zg	න්	n	ඥ්	zn	ල්	l
ළ්	Zl	ය්	Y	ර්	r	ව්	v	ච්	ch
ඣ්	Jh	ස්	S	ශ්	sh	ෂ්	zsh	ත්	txh
ථ්	zth	ද්	dh	ධ්	zdh	හ්	h	ඩ්	d
ඪ්	zdx	ට්	T	ඨ්	zt	බ්	b	භ්	zb
ම්	M	ඹ්	xmb	ප්	p	ඵ්	zp	ඩ	zon
ෆ්	F	ඡ්	ch	ඞ්	zng	ඤ්	cn	ඦ	zjh
ඦ්	jhcn	ඳ	qndh	ඣ	xjhx	ඬ්	zndx	○ං	xonx
○ඃ	zkf	○	W	ෞ	zau	◌ෛ	zaik	○aa	zruu
zwj	qx								

It is easier to use the entire sentence in phonetic English format than having parts of it converted. Therefore, the entire Unicode sentence is converted to phonetic English using an algorithm published in (Punchimudiyanse & Meegama, 2015b) adhering to the above rules. The converted phonetic English SSL sequence is sent to the 3D avatar animation system for live 3D animations.

The Processing performed at the 3D avatar

This research utilizes a 3D avatar and its animation framework, which provides seamless switch over between conversational and fingerspelling signs of SSL (Punchimudiyanse & Meegama, 2015a). It has a posture based sign database, which stores the coordinates of different bones in a posture. The postures of the SSL signs are animated as key frames while the displacement of each bone position in frames in the transition between postures is automatically calculated using a mathematical formula given below.

In this formula, the start value denotes the initial position of a given bone before transition and the end value denotes the final position of a bone after the transition. Number of frames denotes the frames required for the transition to complete. The calculated increment value is the amount of movement required for a bone in direction k per frame in the transition. The block diagram in Figure 3 depicts the

processing, which takes place within the animation framework when a translated SSL sequence is sent to the animation system.
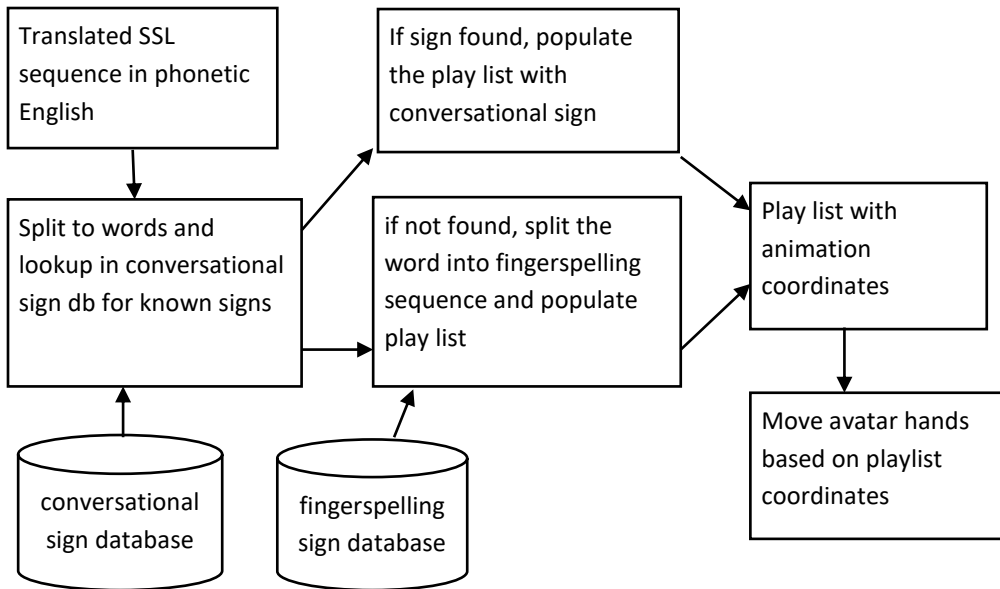
Figure 3: Processing that happens at the 3D avatar animation system

Results and Discussion

The existing sign database of the 3D avatar has over 200 conversational signs related to nouns and verbs while 61 fingerspelling signs are defined to perform fingerspelling. However, the written Sinhala to SSL translation is not popular amongst the SSL community. Therefore, the translation algorithm is tested with a small vocabulary of 50 signs and a word form database of 500+ entries related to the vocabulary. In addition, signs for "finished" and "a lot" are used to denote past tense and plural forms. Since, there are 18 entries per noun and 24 entries per verb in the word form database, it was essential to test the success of the SSL signing technique before scaling up the synonym and word form databases. Therefore, a sample of 100 sentences is prepared to include at least one word form matching the 50 SSL signs. Table 5 indicates 10 translations of the test sample.

Table 5: Translated SSL sequence for 10 sentences in the test sample

Sentence	Converted SSL sequence	Description of word forms
මව බත් කන්නීය	*ම* අම්මා බත් *කනවා* නීය	2 letter synonym and feminine verb
තාත්තා ගෙදර ගියේය	තාත්තා ගෙදර *යනවා ඉවරයි* යේය	Singular past tense verb
අක්කේ පෑන අමතක වුනාද	*අක්කා* කේ පෑන අමතක *වෙනවා ඉවරයි* ද	Word form of noun and past tense question
අයියයි අක්කයි ගෙදරින් බත් කෑවෝය	*අයියා* යි *අක්කා* යි ගෙදර රින් බත් *කනවා ඉවරයි* වෝය	3 nouns with word forms, plural verb form
සමන් බත් කෑවේය	සමන් බත් *කනවා ඉවරයි* වේය	Singular verb form
නුඹ මේස දිහා බැලුවෙහි	නුඹ *මේසය ගොඩක්* දිහා *බලනවා ඉවරයි* හි	2nd person noun, plural noun, past tense 2nd person verb form
තාත්තේ මට උගන්වන්න	තාත්තා තේ මම ට උගන්වනවා න්න	2 nouns with word form, verb form
පියාගේ පෑන් ලස්සනයි	*පි* තාත්තා ගේ පෑන *ගොඩක්* ලස්සන යි	Synonym of noun, plural noun
ගුරුවරයා ගණිතය ඉගැන්වීය	ගුරුවරයා ගණිතය ඉ *උගන්වනවා* වීය	Synonym of verb form
අයියා සහ සමන් මෑණියන්ගේ නිවසට ගියෝය	අයියා සහ සමන් මෑ අම්මා යන් ගේ නි ගෙදර ට යනවා ගොඩක් ඉවරයි	2 synonyms of nouns, past tense plural verb form

The translated SSL sequence of each sentence in the test sample is manually inspected for correct SSL sequence. All the 100 sentences are correctly translated using the written Sinhala to SSL translation algorithm and Table 6 summarizes the translation results.

Table 6: The translation statistics of the algorithm

Description	Count
Sentence count	100
Word count	444
Correct Noun forms	297
Correct verb forms	85
Correctly skipped unknown words	62
Total word forms that are properly translated	444

Translation performance of the algorithm reached 100% because of the small-size of sentence sample and database driven word form replacement. The small test sample enabled the authors to quickly develop and demonstrate the working prototype to SSL user community to motivate them to determine word form to SSL translations instead of authors determining those SSL sequences.

It is observed that the words ending with suffixes "තේ" and "ගේ" to denote a specific word form should originally exist in an input sentence without a space between the base word and suffix to have the correct translation. Otherwise the translator will misinterpret those suffixes individually as their individual meanings "tea" and "house" respectively. For example, the word **"අයියාගේ"** is in correct form, which gives the meaning "brother's" in the written Sinhala translation while "අයියා ගේ" gives the translation sequence as two signs "brother house".

The phonetic English sequence of the SSL translation of the first sentence in Table 5 is "ma amwmaxa batxhw kanavaxa nixiya" and it is sent to the 3D animation system. Figure 4 shows the animation sequence of the 3D avatar.
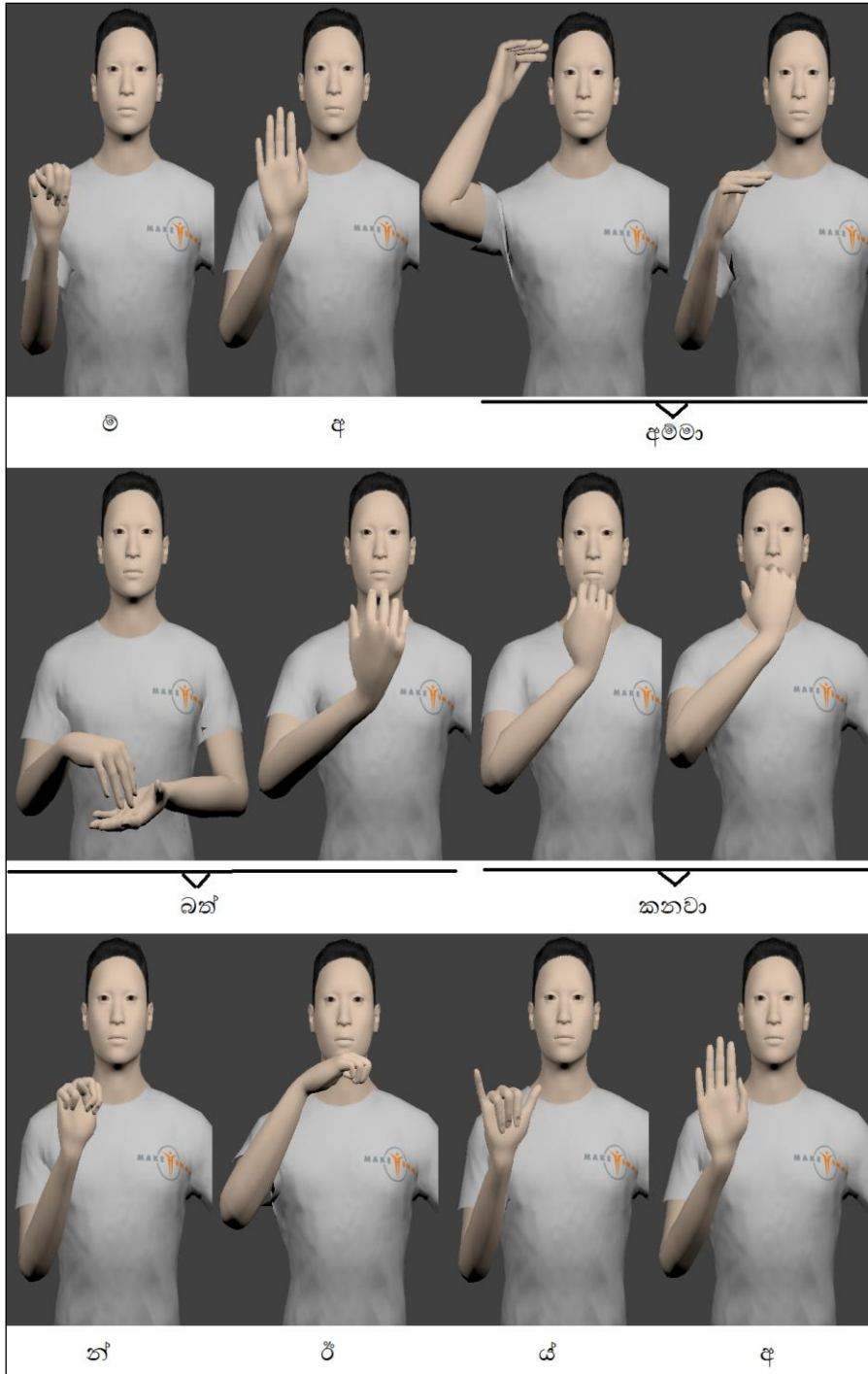
Figure 4 – Animation sequence of *ම අම්මා බත් කනවා නීය* **(** *මව බත් කන්නීය* **)**

Hundred sentences of the test sample were sent through the 3D avatar animation system and the animations were screen captured. Those screen capture videos were shown to three professional interpreters and three SSL teachers without disclosing the transcript. The word form identification rates are given in Table 7.

Table 7: Word form/Sentence Identification rates of 6 SSL users

Interpreter	Identified Word forms (Total 444)	Identified Sentences (Total 100)
1	327	73
2	298	67
3	338	74
4	285	64
5	354	76
6	315	71
Overall %	71.96%	70.83%

Recognition results indicate 70.83% sentence identification rate and 71.96% word form identification rate for the translated SSL sequences out of the 100-sentence test sample. The test users have indicated that the novelty and the lack of regular use of the translation technique have contributed to a lot of word form identification errors. Moreover, they indicated that the animation of the fingerspelling character එ needs to be slowed down in animations in order to properly identify suffixes such as "ගේ" and "තේ". Furthermore, they have suggested to show all the SSL signs related to a word form in a single motion to indicate that they are part of a single word form. For example, the signs "පි","තාත්තා", "ට" in the word තාත්තාට has to be signed in a continuous motion without showing space character in between signs.

Test users are optimistic that the sign identification rate of the system will improve over time when this technique is in regular use.

Conclusion

This study highlights a small-scale implementation to build a computerized translator for emulating the translation technique of written Sinhala to SSL. A SSL sequencing algorithm with synonym and word form databases provides the backbone of the system, which is tested with a small sign vocabulary to leave room for expansion.

Moreover, translated SSL sequence is generated in the format compatible with the 3D avatar system already developed by us.

Our observations indicate that this technique is less popular among the SSL community due to lack of training and the non-existence of a common word form to SSL translation vocabulary. Therefore, we propose to expand the word form and synonym database of the system with the help of the SSL community before performing a comprehensive analysis on computerized translation performance. In the computer science perspective, many word forms that exist for a Sinhala word will expand the word form database into millions of entries. This will make searching for entries in the word form database slower than expected. Therefore, it is necessary to find patterns in word forms to reduce the size of word form database or use indexed searching techniques in future after gathering a large number of SSL sequences for different Sinhala word forms with the support of the SSL user community.

Acknowledgement

References

Almeida, I, Coheur, L., & Candeias, S. (2015). Coupling Natural Language Processing and Animation Synthesis in Portuguese Sign Language Translation, *Proceedings of the Workshop on Vision and Language (VL'15)*, (pp. 94–103), Lisbon, Portugal: Association for Computational Linguistics.

Awad, C., Courty, N., Duarte, K., Le Naour, T., & Gibet, S. (2009). A Combined Semantic and Motion Capture Database for Real-Time Sign Language Synthesis. *Proceedings of the 9th International Conference on Intelligent Virtual Agent (IVA 2009),*(pp. 432-438), Amsterdam, Netherlands.<hal-00493426>

Chiu, Y. H., Wu, C. H., Su, H. Y., & Cheng, C. J. (2007). Joint optimization of word alignment and epenthesis generation for Chinese to Taiwanese sign synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1), 28-39.

Chuang, Z. J., Wu, C. H., & Chen, W. S. (2006). Movement epenthesis generation using NURBS-based spatial interpolation. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(11), 1313-1323.

Cox, S., Lincoln, M., Tryggvason, J., Nakisa, M., Wells, M., Tutt, M., & Abbott, S. (2002). Tessa, a system to aid communication with deaf people. *Proceedings of the 5ᵗʰ international ACM conference on Assistive technologies,* (pp. 205-212), Edinburgh, United Kingdom: ACM.

Davidson, M. J. (2006). PAULA: A Computer-Based Sign Language Tutor for Hearing Adults, *Proceedings of the 8th International Conference on Intelligent Tutoring Systems (ITS 2006): Teaching with Agents, Robots and NLP*, Jhongli, Taiwan.

De Martino, J. M., Costa, P. D. P., Benetti, A., Rosa, L. A., Kumada, K. M. O., & Silva, I. R. (2016). Building a brazilian portuguese-brazilian sign language parallel corpus using motion capture data. *Proceedings of Workshop of Corpora and Tools for Processing Corpora,* (pp. 56-63), Tomar, Portugal.

Dias, N. S. & Dias, S., (1996). සිංහල සංඥා අක්ෂර මිශ්‍ර හෝඩිය, National Institute of Education Press, Sri Lanka

Efthimiou, E., Fotinea, S. E., Vogler, C., Hanke, T., Glauert, J., Bowden, R., ... & Segouat, J. (2009). Sign language recognition, generation, and modelling: a research effort with applications in deaf communication. *Proceedings of International Conference on Universal Access in Human-Computer Interaction* (pp. 21-30). Springer: Berlin Heidelberg.

Elliott, R., Glauert, J., Kennaway, R., & Parsons, K. J. (2001). *D5-2: sigml definition. working document*, University of East Anglia, UK, ViSiCAST Project.

Elliott, R., Glauert, J. R., Kennaway, J. R., Marshall, I., & Safar, E. (2008). Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society*, *6*(4), 375-391.

eSIGN project (2002). Retrieved from : http://www.sign-lang.uni-hamburg.de/esign/

European Sign Language Center. (2012), *Learn sign language by translating written text to signs "spread the sign"*, Retrieved from https://www.spreadthesign.com/us/

Huenerfauth, M. (2006). *Generating American Sign Language classifier predicates for English-to-ASL machine translation,* (Doctoral dissertation, University of Pennsylvania). Retrieved from

https://pdfs.semanticscholar.org/8990/82e3d8abe916f12f3cea6d71f830ba4a486e.pdf.

Huenerfauth, M. (2008). Generating American Sign Language animation: overcoming misconceptions and technical challenges, *Universal Access in the Information Societ*y 6(04):419-434. doi:10.1007/s10209-007-0095-7

Kaneko, H., Hamaguchi, N., Doke M., & Inoue, S.(2010). Sign language animation using TVML, *Proceedings of VRCAI '10*, (pp. 289-292), Seoul, Korea.

Kouremenos, D., Fotinea, S. E., Efthimiou, E., & Ntalianis, K. (2010). A prototype Greek text to Greek Sign Language conversion system. *Behaviour & Information Technology*, *29*(5), 467-481.

Kulaveerasingam, N., Wellage, S., Samarawickrama, H. M. P., Perera, W. M. C., & Yasas, J. (2014). *The Rhythm of Silence - Gesture Based Intercommunication Platform for Hearing-impaired People (Nihanda Ridma).* Retrieved from http://cdap.sliit.lk/bitstream/123456789/279/1/ Nihanda_Ridma.pdf

Liyanapathirana, L. P. M. B. (2014). Text to Sign Language Interpreter for Sinhala.

Marshall, I., & Safar, E. (2005). Grammar development for sign language avatar-based synthesis. *Proceedings of HCII* , (pp. 1-10).

National Institute of Education (NIE SL) (1989). සිංහල ලේඛන රීතිය, Maharagma, Sri Lanka: NIE press, Sri Lanka.

Ong, S. C., & Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. IEEE transactions on pattern analysis and machine intelligence, 27(6), 873-891.

Othman, A., & Jemni, M. (2011). Statistical sign language machine translation: from English written text to American sign language gloss. arXiv preprint arXiv:1112.0168.

Punchimudiyanse, M. & Meegama, R. G. N. (2015). 3D Signing Avatar for Sinhala Sign Language, *In Proceedings of IEEE International Conference on Industrial and Information Systems (ICIIS 2015)*, 290-295. doi:10.1109/ICIINFS.2015.7399026

Punchimudiyanse, M. & Meegama, R. G. N. (2015). Unicode Sinhala and phonetic English bi-directional conversion for Sinhala speech recognizer, *In Proceedings of IEEE International Conference on Industrial and Information Systems (ICIIS 2015)*, 296-301. doi:10.1109/ICIINFS.2015.7399027

Siyambalagoda, W., Weerasena, V., Susantha, K. K. B, Dias, N. S., Rajapaksha, N. N, Hewapathirana, M. C. B., Premasiri, M. R., Hareed, M. S. H. M, Gunasekara, V., Hettiarachchi, H. R. S. K., Sinhanatha, W. V. (Eds) (2004). *සිංහල කියවීම් පොත සඳහා සංඥා වදන් මාලාව [SSL vocabulary for Sinhala reading book I]*, Primary and Special Education Division, National Institute of Education, Sri Lanka.

Solina, F., Krapez, S., Jaklic, A., & Komac, V. (2000). Multimedia dictionary and synthesis of sign language. *Design and management of multimedia information systems*, 268-281.

Stone, A. (2007). *An introduction to Sri Lankan Sign Language*. Matara, Sri Lanka : Karunaratne & Sons, Sri Lanka.

T2VLab, NHK Science and Technical Research Laboratories (2011). *TVML Language Specification Version 3.0*, Retrieved from http://wjis76df.sakura.ne.jp/t2vlab/spec/tvmlspec4t2v/en/

Unicode.org, (2016). *The Unicode Standard, Version 9.0* - Sinhala Range : 0D80-0DFF.