English to Indonesian Transliteration to Support English Pronunciation Practice

Amalia Zahra, Julie Carson-Berndsen

CNGL, School of Computer Science and Informatics, University College Dublin Belfield, Dublin 4, Ireland amalia.zahra@ucdconnect.ie, julie.berndsen@ucd.ie

Abstract

The work presented in this paper explores the use of Indonesian transliteration to support English pronunciation practice. It is mainly aimed for Indonesian speakers who have no or minimum English language skills. The approach implemented combines a rule-based and a statistical method. The rules of English-Phone-to-Indonesian-Grapheme mapping are implemented with a Finite State Transducer (FST), followed by a statistical method which is a grapheme-based trigram language model. The Indonesian transliteration generated was used as a means to support the learners where their speech were then recorded. The speech recordings have been evaluated by 19 participants: 8 English native and 11 non-native speakers. The results show that the transliteration positively contributes to the improvement of their English pronunciation.

Keywords: English to Indonesian transliteration, pronunciation practice, Finite State Transducers, language models

1. Introduction

Transliteration is the task of converting words in a source language to phonetically equivalent words in a target language, which means the original pronunciation in the source language is preserved. For instance, the word "language" may be transliterated to "lengwec" in Indonesian. Due to the fact that each language has different phonetic inventories, the challenge is to find substitutions for the phones that exist in the source language but do not in the target language or the phones that are valid in some position within a word in English but invalid in Indonesian for the same position. For instance, the phone /jh/ is never pronounced at the end of a word in Indonesian which might make Indonesian speakers mispronounce it badly that could lead to confuse other people. The substitutions are selected by investigating how similar they are to the substituted phone in terms of place and manner of articulation.

Transliteration has been widely used in various areas, such as Machine Translation (MT), Cross Language Information Retrieval (CLIR), Information Extraction (IE), Textto-Speech (TTS), etc. In those areas, it is commonly used to transliterate proper names across languages, such as names of people, places, companies, etc. Several languages have been included in the studies of transliteration. Ali and Ijaz (2009) implemented pronunciation-based transliteration from English to Urdu. Their work applied rule-based mapping, syllabification, and urduization rules. An English to Tamil transliteration system (Vijayanand et al., 2009) was developed to handle named entities in order to ensure the quality of English-Tamil Machine Translation. Automatic transliteration has also been investigated for English to Hindi, Tamil, Kannada, Russian, Chinese, Korean Hangul, Japanese Katakana, and Japanese Kanji (NEWS 2009 Machine Transliteration Shared Task) (Oh et al., 2009), English to Arabic (Sherif and Kondrak, 2007), English to Japanese (Finch and Sumita, 2008) and many more. The methods applied have varied: a dynamic programming algorithm and a finite state transducer (Sherif and Kondrak, 2007), phrase-based machine transliteration (Finch and Sumita, 2008), target-language-grapheme and target-language-grapheme-and-phoneme based transliteration model (Oh et al., 2009), rule-based approach (Vijayanand et al., 2009), the implementation of Statistical Machine Translation (SMT) (Antony et al., 2010), and many more. The above studies were carried out mainly to support the tasks of Machine Translation and Cross Language Information Retrieval.

The work presented here has a different goal to those presented above. To be precise, this work aims to support English pronunciation learning for Indonesian speakers who cannot speak English or only know very little about English. As part of learning a new language, pronunciation is as important as grammar and vocabulary. Words have to be pronounced correctly in order to be understood by the other person. This proves difficult for Indonesian people who are not familiar with English. This is due to different phonetic inventories and to the fact that different pronunciation rules apply in English and Indonesian. For instance, the word "day" would be pronounced as /d ay/ (like in the word "die") by Indonesian speakers as that is the rule of how to pronounce this combination of graphemes in Indonesian. Hence, the pronunciation of the word "day" should be transformed into the form that will be understood by Indonesian speakers, which is "dey" in this case. This work investigates whether an English to Indonesian transliteration system is helpful in terms of supporting Indonesian speakers to learn English pronunciation and whether it helps them enhance their ability to pronounce English words.

The remainder of this paper is structured as follows. Section 2 explains about English to Indonesian transliteration which is a brief description of the approach applied and the mapping involved. The overall system architecture is explained in Section 3. The experimental setup and the subjective evaluation are described in Sections 4 and 5 respectively, followed by some results and analysis which are presented in Section 6. Finally, some conclusions are drawn in Section 7 along with some closing remarks.

2. English to Indonesian Transliteration

Two techniques have been applied in the English to Indonesian transliteration system: a Finite State Transducer (FST) and a trigram Language Model (LM). In contrast to the substring-based FST implemented in (Sherif and Kondrak, 2007), the FST applied in this approach consists of the mapping from individual English phones to individual Indonesian graphemes. The mapping can be performed from the source phones to the target graphemes directly because almost all of Indonesian graphemes have one-to-one mapping to their phones and they use the Roman alphabets, as does English. English has around 45 phones (TIMIT (Garofolo et al., 1993)) while Indonesian has only 33 phones (Zahra et al., 2009); thus, there are English phones that do not exist in Indonesian. These phones are replaced by their closest counterparts by considering place and manner of articulation. On the other hand, those Indonesian phones that do not exist in English are ignored since they are not used in the forward transliteration (English to Indonesian). Table 1 outlines the mapping between English phones and Indonesian graphemes covered by the FST.

No	English	Indonesian	No	English	Indonesian
	Phones	Graphemes		Phones	Graphemes
1	/aa/	o, a	24	/iy/	i
2	/ae/	e	25	/jh/	j, c
3	/ah/	а	26	/k/	k
4	/ao/	о	27	/1/	1
5	/aw/	aw	28	/m/	m
6	/ax/	Е	29	/n/	n
7	/axr/	Er	30	/ng/	ng
8	/ay/	ay	31	/ow/	о
9	/b/	b	32	/oy/	oy
10	/ch/	с	33	/p/	р
11	/d/	d	34	/r/	r
12	/dh/	d	35	/s/	s
13	/eh/	e	36	/sh/	sy
14	/el/	El	37	/t/	t
15	/em/	Em	38	/th/	th
16	/en/	En	39	/uh/	u
17	/er/	Er	40	/uw/	uw
18	/ey/	ey	41	/v/	f, v
19	/f/	f	42	/w/	w
20	/g/	g	43	/y/	у
21	/hh/	h	44	/z/	z
22	/ih/	i	45	/zh/	sy
23	/ix/	E, i			

Table 1: English phones to Indonesian graphemes mapping

In Indonesian, the letter "e" has two pronunciations, either ϵ /or / ϵ / or / ϵ / (IPA (Ladefoged, 1990)). They are distinguished by using the capital letter "E". This is included in the mapping, as shown in Table 1.

3. The System Architecture

As mentioned in Section 2, English phones are mapped directly to their corresponding Indonesian graphemes, which is implemented with an FST. In the case of multiple outputs or alternatives, a grapheme-based trigram language model is applied to obtain the final transliteration. Figure 1 shows the overall architecture of the English to Indonesian transliteration system.

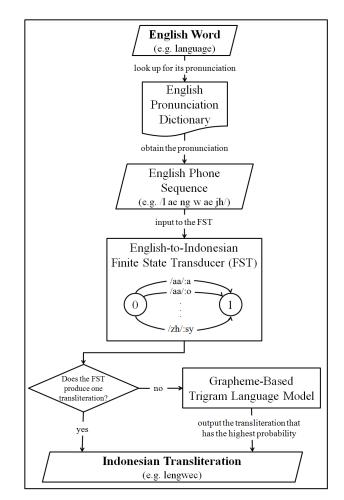


Figure 1: The architecture of the English to Indonesian transliteration system.

Given an English word (e.g. "language"), the system firstly looks for its pronunciation from the English pronunciation dictionary. The phone sequence is then passed to the FST to obtain one or several possible Indonesian grapheme sequences. If it produces one grapheme sequence, then the sequence is immediately prompted to users. Otherwise, a grapheme-based trigram Language Model is applied to select the final grapheme sequence, namely the one with the highest probability (e.g. "lengwec"). The data used to train the language model comes from the manually transliterated words from the full TIMIT dictionary, where all words for testing have been removed. The number of words for training the language model is 6133 words while it is 96 words used for testing (i.e. used by the Indonesian speakers for pronunciation learning). The training was performed by using SRILM toolkit (Stolcke, 2002).

4. Experimental Setup

As mentioned earlier, the transliteration output is used to support Indonesian speakers learn English pronunciation. In order to investigate whether this approach is useful, a learning scenario has been developed, which involves three steps. Firstly, the user is asked to read the English sentence without any help (attempt-1). Secondly, the user is allowed to listen to a reference speech file as many times as he/she wants, and is then asked to read the sentence again (attempt-2). Finally, an Indonesian transliteration is prompted to the user so that he/she can read it, and then asked to say the same sentence (attempt-3).

Thus, for each English sentence, there are three recording files for each speaker. Five Indonesian speakers took part in the experiment, two have no English knowledge and three have minimum English. The first two have no knowledge of other languages except Indonesian and their own traditional languages. This is the same for the remaining three except they know English in limited amount. The sentences read by the speakers were common phrases, such as "See you in the morning.", "Can I see the menu, please?", etc. In total, there are 45 sets of recordings. Each set consists of three speech files representing three attempts outlined above.

5. Subjective Evaluation

Once the speech recordings were collected, the next step was to perform subjective evaluation on all 45 sets of recordings. For each set, evaluators were asked to select which of the three speech files represents the best pronunciation. The three speech files for each set were presented in a random order. Figure 2 shows a snapshot of the subjective evaluation page.

oronunciation.	•
Test 1	
Sentence 1: "D	o you have any rooms free?"
Speech A:	
© Speech B:	
© Speech C:	

Figure 2: Snapshot of the subjective evaluation page.

The evaluation was performed by 19 participants: 8 English native and 11 non-native speakers. The reason for selecting regular speakers as the evaluators, both native and non-native, was based on why people learn language pronunciation in the first place, which is for the purpose of intelligibility. If a person perceives correctly (i.e. understands) what the other person is saying, then the communication between them is more likely to succeed.

6. Results and Analysis

The opinions of all evaluators, both native and non-native, were processed by computing the average percentage of how many times they selected the speech recordings coming from the first, second, and third attempt as having the best pronunciations. The results are divided into three categories, for the speakers who have no English knowledge ("No English"), minimum English knowledge ("Minimum English"), and all speakers ("All"), which are shown in Figure 3.

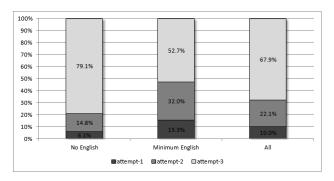


Figure 3: The percentages of how many times evaluators select the speech recordings from attempt-1, -2, and -3 as having the best pronunciations.

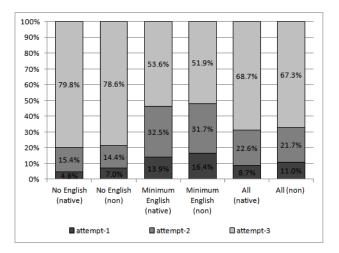


Figure 4: The percentages of how many times English native (native) and non-native (non) evaluators select the speech recordings from attempt-1, -2, and -3 as having the best pronunciations.

Figure 3 shows that the speakers with no knowledge of English find the Indonesian transliteration most useful to improve their English pronunciation as 79.1% of their attempt-3 recordings are selected as having the best pronunciation by the evaluators. This has significant difference to the 14.8% of their attempt-2 recordings selected by the evaluators. Apparently, listening to the reference speech files (the condition for attempt-2) is not enough to improve their English pronunciation, even though the speakers are allowed to do so as many times as they want.

For the speakers who have minimum English, the percentages of their attempt-2 and attempt-3 speech recordings being selected as having the best pronunciation by the respondents are not significantly different. The selected recordings from attempt-2 and attempt-3 are 52.7% and 32%, respectively. However, the Indonesian transliteration can still be considered useful as there is a 20.7% absolute increase between the attempt-2 and attempt-3. Overall, 67.9% and 22.1% of the speech recordings from attempt-2 and attempt-3, respectively, are selected by the evaluators. In addition to that, it is interesting to see if there is significant difference in opinions between English native and non-native evaluators. Figure 4 shows the same results as those in Figure 3 but separated between the opinions of English native evaluators and those of non-native.

As can be seen in Figure 4, there is no significant difference between opinions of native and non-native evaluators either for "No English", "Minimum English", or "All" category. This means that a strong agreement exists between the two types of evaluators.

7. Conclusions and Closing Remarks

This work aims to explore a transliteration system to assist Indonesian speakers learn English pronunciation and investigate its level of usefulness to users. The percentage differences between attempt-2 and attempt-3 presented in Figure 3 and 4 and the existence of a strong agreement between the English native and non-native evaluators demonstrate that the Indonesian transliteration system is a useful tool to help Indonesian speakers who have no or minimum English to learn English pronunciation. One might think the reason that most of the best pronunciations are obtained from attempt-3 is due to the condition where the speakers were allowed to listen to the reference speeh as many times as they want. However, they were also allowed to practice their pronunciations before being recorded for attempt-2, yet the Indonesian transliteration provided for attempt-3 undoubtly contributes to the positive improvement on their pronunciation learning.

Two things that have not yet been taken into account in the experiments due to time constraints are collecting speech from more speakers and randomizing the order of condition 2 and 3 for different phrases. This would have made the conclusion stronger if the results obtained follow the same pattern as having been presented in this paper.

In the future, this work will be integrated with the Computer-Aided Pronunciation Learning (CAPL) system developed by our research group (Cabral et al., 2012). Since stress plays an important role in English, it might presumably be beneficial to apply some prosodic transliteration (e.g. using uppercase for stressed syllables) as an extra aid for the learners. Furthermore, it will be interesting to investigate whether this approach is also useful for Indonesian speakers with medium to advanced English skills to further improve their English pronunciation.

8. Acknowledgements

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at University College Dublin (UCD). The opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Science Foundation Ireland.

9. References

- A. R. Ali and M. Ijaz. 2009. English to urdu transliteration system. In *Proceedings of the Conference of Language and Technology, Lahore, Pakistan.*
- P. J. Antony, V. P. Ajith, and K. P. Soman. 2010. Statistical method for english to kannada transliteration. In *Information Processing and Management - International Conference on Recent Trends in Business Administration and Information Processing, BAIP*, volume 70 of *Communications in Computer and Information Science*, pages 356–362. Springer.
- J. P. Cabral, M. Kane, Z. Ahmed, M. Abou-Zleikha, É. Székely, A. Zahra, K. U. Ogbureke, P. Cahill, J. Carson-Berndsen, and S. Schlögl. 2012. Rapidly testing the interraction model of a pronunciation training system via wizard-of-oz. In *Proceedings of the 8th international conference on Language Resources and Evaluation LREC, Istanbul, Turkey.*
- A. Finch and E. Sumita. 2008. Phrase-based machine transliteration. In Proceedings of the Workshop on Technologies and Corpora for Asia-Pacific Speech Translation (TCAST), pages 13–18.
- J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren. 1993. DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. National Institute of Standards and Technology, NISTIR 4930.
- P. Ladefoged. 1990. The revised international phonetic alphabet. *Journal of the International Phonetic Association*, 19(2):67–80.
- J. H. Oh, K. Uchimoto, and K. Torisawa. 2009. Machine transliteration using target-language grapheme and phoneme: Multi-engine transliteration approach. In *Proceedings of the 2009 Named Entities Workshop: Shared Task on Transliteration (NEWS 2009)*, pages 36–39. The Association for Computational Linguistics.
- T. Sherif and G. Kondrak. 2007. Substring-based transliteration. In ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic. The Association for Computational Linguistics.
- A. Stolcke. 2002. SRILM an extensible language modeling toolkit. In *Proceedings of the 7th International Conference on Spoken Language Processing, ICSLP2002 -INTERSPEECH, Denver, Colorado, USA.* ISCA.
- K. Vijayanand, I. R. Babu, and P. Sandiran. 2009. Testing and performance evaluation of machine transliteration system for tamil language. In *Proceedings of the* 2009 Named Entities Workshop, pages 48–51. The Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP).
- A. Zahra, S. Baskoro, and M. Adriani. 2009. Building a pronunciation dictionary for indonesian speech recognition system. In Proceedings of the Workshop on Technologies and Corpora for Asia-Pacific Speech Translation (TCAST), Singapore.