論文 / 著書情報
Article / Book Information

Title	Audio-visual Speech Recognition Using Lip Information Extracted from Side-face Images
Authors	Koji Iwano, Tomoaki Yoshinaga, Satoshi Tamura, Sadaoki Furui
Citation	EURASIP Journal on Audio, Speech, and Music Processing, Vol. 2007, No. , pp. Article ID 64506
Pub. date	2007, 3
Creative Commons	See next page.

License

Hindawi Publishing Corporation
EURASIP Journal on Audio, Speech, and Music Processing
Volume 2008, Article ID 573832, 7 pages
doi:10.1155/2008/573832

Research Article

Language Model Adaptation Using Machine-Translated Text for Resource-Deficient Languages

Arnar Thor Jensson, Koji Iwano, and Sadaoki Furui

Department of Computer Science, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan

Correspondence should be addressed to Arnar Thor Jensson, arnar@furui.cs.titech.ac.jp

Received 30 April 2008; Revised 25 July 2008; Accepted 29 October 2008

Recommended by Martin Bouchard

Text corpus size is an important issue when building a language model (LM). This is a particularly important issue for languages where little data is available. This paper introduces an LM adaptation technique to improve an LM built using a small amount of task-dependent text with the help of a machine-translated text corpus. Icelandic speech recognition experiments were performed using data, machine translated (MT) from English to Icelandic on a word-by-word and sentence-by-sentence basis. LM interpolation using the baseline LM and an LM built from either word-by-word or sentence-by-sentence translated text reduced the word error rate significantly when manually obtained utterances used as a baseline were very sparse.

1. INTRODUCTION

The state-of-the-art speech recognition has advanced greatly for several languages [1]. Extensive databases both acoustical and text have been collected in those languages in order to develop the speech recognition systems. Collection of large databases requires both time and resources for each of the target language. More than 6000 living languages are spoken in the world today. Developing a speech recognition system for each of these languages seems unimaginable, but since one language can quickly gain political and economical importance a quick solution toward developing a speech recognition system is important.

Since data, for the purpose of developing a speech recognition system, is sparse or nonexisting for resource-deficient languages, it may be possible to use data from the other resource-rich languages, especially when available target language sentences are limited which often occurs when developing prototype systems.

Development of speech recognizers for resource-deficient languages using spoken utterances in a different language has already been reported in [2], where phonemes are identified in several different languages and used to create or aid an acoustic model for the target language. Text for creating the language model (LM) is on the other hand assumed to exist in a large quantity and therefore sparseness of text is not addressed in [2].

Statistical language modeling is well known to be very important in large vocabulary speech recognition but creating a robust language model typically requires a large amount of training text. Therefore it is difficult to create a statistical LM for resource deficient languages. In our case, we would like to build an Icelandic speech recognition dialogue system in the weather information domain. Since Icelandic is a resource deficient language there is no large text data available for building a statistical LM, especially for spontaneous speech.

Methods have been proposed in the literature to improve statistical language modeling using machine-translated (MT) text from another source language [3, 4]. A cross-lingual information retrieval method is used to aid an LM in different language in [3]. News stories are translated from a resource-*rich* language to a resource-*sparse* language using a statistical MT system trained on a sentence-aligned corpus in order to improve the LM used to recognize similar or the same story in the resource-*sparse* language. Another method described in [4] uses ideas from latent semantic analysis for cross lingual modeling to develop a single low-dimensional representation shared by words and documents in both languages. It uses automatic speech

Table 1: Datasets.

Corpus set	Sentences	Words	Unique words
ST	1500	8591	805
SD	300	1870	342
Eval	660	3767	554

recognition transcripts and aligns each with the same or similar story in another language. Using this parallel corpus a statistical MT system is trained. The MT system is then used to translate a text in order to aid the LM used to recognize the same or similar story in the original language. LM adaptation with target task machine-translated text is addressed in [5] but without speech recognition experiments. A system that uses an automatic speech recognition system for human translators is improved in [6] by using a statistical machine translation of the source text. It assumes that the content of the text translated is the same as in the target text recognized. The above mentioned systems all use statistical machine translation (MT) often expensive to obtain and unavailable for resource-deficient languages.

MT methods other than statistical MT are also available, such as rule based MT systems. A rule based MT system can be based on a word-by-word (WBW) translation or sentence-by-sentence (SBS) translation. WBW translation only requires a dictionary, already available for many language pairs, whereas rule based SBS MT needs more extensive rules and therefore more expensive to obtain. The WBW approach is expected to be successful only for closely grammatical related languages. In this paper, we investigate the effectiveness of WBW and SBS translation methods and show the amount of data for the resource-deficient language required to par these methods.

In Section 2, we explain the method for adapting language models. Section 3 explains the experimental corpora. Section 4 explains the experimental setups. Experimental results are reported in Section 5 followed by a discussion in Sections 6, and 7 concludes the paper.

2. ADAPTATION METHOD

Our method involves adapting a task-dependent LM that is created from a sparse amount of text using a large translated text (TRT), where TRT denotes the machine translation of the rich corpus (RT), preferably in the same domain area as the task. This involves two steps shown graphically in Figure 1. First of all the sparse text is split into two, a training text corpus (ST) and a development text corpus (SD). A language model LM1 is created from ST, and LM2 from TRT. The TRT can either be obtained from SBS or WBW translation. The SD set is used to optimize the weight (λ) used in Step 2. Step 2 involves interpolating LM1 and LM2 linearly using the following equation:

$$P_{\text{comb}}(\omega_i \mid h) = \lambda \cdot P_1(\omega_i \mid h) + (1-\lambda)P_2(\omega_i \mid h), \quad (1)$$

where h is the history. P_1 is the probability from LM1 and P_2 is the probability from LM2.

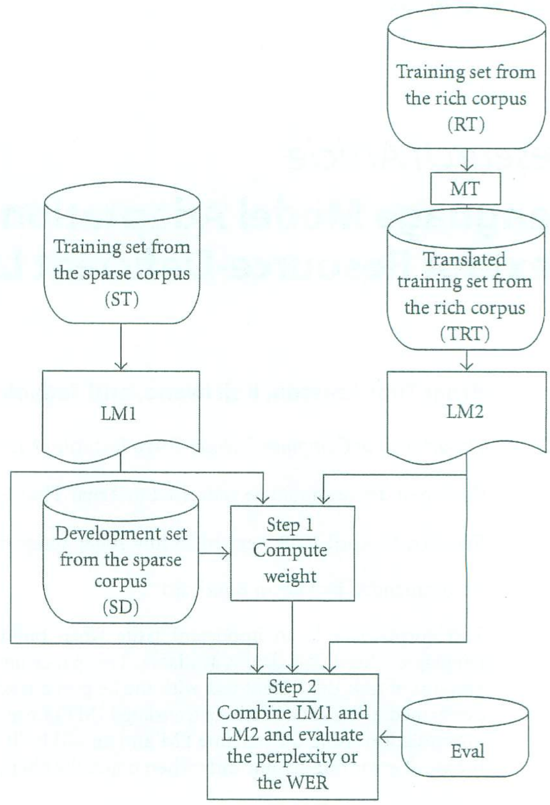

Figure 1: Data diagram.

The final perplexity or word error rate (WER) value is calculated using an evaluation text set or speech evaluation set ($Eval$) which is disjoint from all other datasets.

3. EXPERIMENTAL CORPORA

3.1. Experimental data: LM

The weather information domain was chosen for the Icelandic experiments and translation from English (*rich*) to Icelandic (*sparse*) using WBW and SBS. For the experiments, the Jupiter corpus [7] was used. It consists of unique sentences gathered from actual users' utterances. A set of 2460 sentences were manually translated from English to Icelandic and split into ST, SD, and $Eval$ sets as shown in Table 1. 63116 sentences were used as RT.

A unique word list was made out of the Jupiter corpus, and was machine translated using [8] in order to create a dictionary. This MT is a rule-based system. The dictionary consists of one-to-one mapping, that is, an original English word has only one Icelandic translation. The word translation can consist of zero (unable to translate), one, or multiple words. Multiple words occur in the case when a word in English cannot be described in one word in Icelandic such that the English word "today" translates to the Icelandic words "dag." An English word is usually translated to one Icelandic word only.

Table 2: Translated datasets.

Corpus set	Sentences	Words	Unique words
TRT_{WBW}	62962	440347	3396
TRT_{SBS}	62996	406814	7312

Table 3: BLEU evaluation of the *SBS* and the *WBW* machine translators.

Translation method	BLEU				
	1-gram	2-gram	3-gram	4-gram	Average
WBW	0.47	0.28	0.19	0.15	0.27
SBS	0.58	0.42	0.32	0.26	0.39

Table 4: Icelandic phonemes in IPA format.

Vovel	/ ɪ, i, ɛ, a, ɤ, œ, u, ɔ, au, ou, ei, ai, œy /
Consonant	/ p, pʰ, t, tʰ, c, cʰ, f, v, ð, s, j, ç, ɣ, m, n, l, r /

The dictionary was then used to translate *RT* WBW into TRT_{WBW}. Another translation TRT_{SBS} was created by SBS machine translation using [8]. Names of places were identified and then replaced randomly with Icelandic place names for both TRT_{WBW} and TRT_{SBS}, since the task is in the weather information domain. Table 2 shows some attributes of the WBW and SBS translated Jupiter texts. The reason why the number of sentences in Table 2 does not match the number of sentences found in the *RT* set is because of empty translations. The reason why the unique words in Table 2 are more than double for TRT_{SBS} compared to TRT_{WBW} is because Icelandic is a highly inflected language and the SBS translation system can cope with those kinds of words as well as word tenses and words articles to some extent whereas the WBW translation system copes poorly.

A 1-gram, 2-gram, 3-gram, and 4-gram translation evaluation using BLEU [9] was performed on 100 sentences created from both the SBS and the WBW machine translators, using two human references. Table 3 shows the BLEU evaluation results. The SBS machine translation outbeats the simple WBW translation as expected. It is a known fact that even human translators do not get full mark (1.0) using the BLEU evaluation [9]. The evaluation still results in 0.15 and 0.26 for WBW and SBS, respectively, using 4-gram evaluation.

3.2. *Experimental data: acoustic model*

A biphonetically balanced (PB) Icelandic text corpus was used to create an acoustic training corpus. A text-to-phoneme translation dictionary was created for this purpose based on [10] using 257 pronunciation rules. The whole set of 30 Icelandic phonemes used to create the corpus, consisting of 13 vowels and 17 consonants, are listed in IPA format in Table 4.

Some attributes of the PB corpus are given in Table 5. The acoustic training corpus was then recorded in a clean environment to minimize external noise. Table 6 describes some attributes of the acoustic training corpus.

Table 5: Some attributes of the phonetically balanced Icelandic text corpus.

Attribute	Text corpus
No. of sentences	290
No. of words	1375
No. of phones	8407
PB unit	Biphone
No. of unique PB units	916
Average no. of words/sentence	4.7
Average no. of phones/word	6.1

Table 6: Some attributes of the Icelandic acoustic training corpus.

Attribute	Acoustic corpus
No. of male speakers	13
No. of female speakers	7
Time (hours)	3.8

Table 7: Some attributes of the Icelandic evaluation speech corpus.

Attribute	Evaluation speech corpus
No. of utterances	4000
No. of male speakers	10
No. of female speakers	10
Time (hours)	2.0

Table 8: Experimental setup.

Experiment no.	*TRT* corpus	Vocabulary
Experiment 1	None	V_{ST}
Experiment 2	None	$V_{ST} + V_{TRT_{WBW}}$
Experiment 3	TRT_{WBW}	V_{ST}
Experiment 4	TRT_{WBW}	$V_{ST} + V_{TRT_{WBW}}$
Experiment 5	None	$V_{ST} + V_{TRT_{SBS}}$
Experiment 6	TRT_{SBS}	V_{ST}
Experiment 7	TRT_{SBS}	$V_{ST} + V_{TRT_{SBS}}$
Experiment 8	$TRT_{WBW} + TRT_{SBS}$	$V_{ST} + V_{TRT_{WBW}} + V_{TRT_{SBS}}$

25-dimensional feature vectors consisting of 12 MFCCs, their delta, and a delta energy were used to train gender-independent acoustic model. Phones were represented as context-dependent, 3-state, left-to-right hidden Markov models (HMMs). The HMM states were clustered by a phonetic decision tree. The number of leaves was 1000. Each state of the HMMs was modeled by 16 Gaussian mixtures. No special tone information was incorporated. HTK [11] version 3.2 was used to train the acoustic model.

3.3. *Evaluation speech corpus*

An evaluation corpus was recorded using sentences from the previously explained *Eval* set. There were 660 sentences in total but divided into sets of 220 sentences for each speaker, overlapping every 110 sentences. The final speech evaluation corpus was stripped down to 200 sentences for

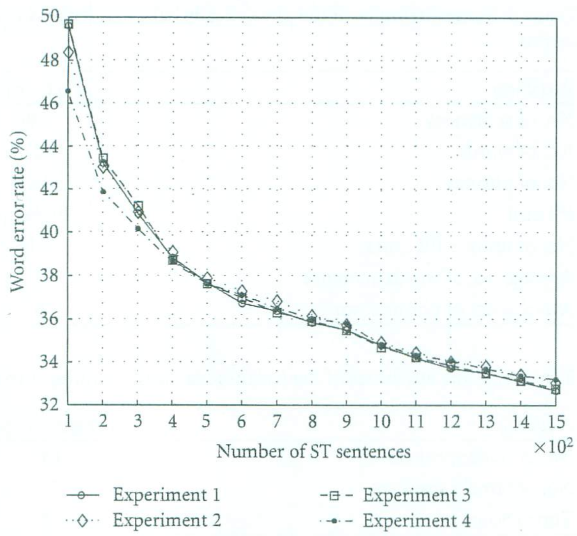

Figure 2: Word error rate results using the *baseline* from Experiment 1 and the interpolated WBW machine-translated results from Experiment 2, Experiment 3, and Experiment 4.

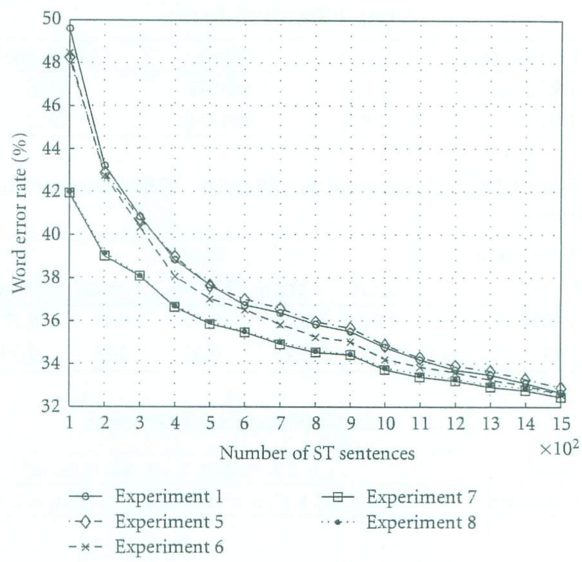

Figure 3: Word error rate results using the *baseline* from Experiment 1 and the interpolated SBS machine-translated results from Experiment 5, Experiment 6, Experiment 7, and Experiment 8.

each speaker since several utterances were deemed unusable. Some attributes of the corpus are presented in Table 7. None of the speakers in the evaluation speech corpus is included in the acoustic training corpus described in Section 3.2.

4. EXPERIMENTAL SETUP

In total, eight different experiments were performed. The experimental setup can be viewed in Table 8. Experiment 1 used no translation and its vocabulary consisted only from the unique words found in the ST set, creating V_{ST}, and is therefore considered as the *baseline*. Experiments 2 to 4 used WBW machine-translated data. Experiment 2 used no TRT corpus but used the unique words found in TRT_{WBW}, creating the vocabulary $V_{TRT_{WBW}}$. This was done in order to find the impact of including only WBW translated vocabulary. Experiment 3 used the WBW machine-translated corpus along with the V_{ST} vocabulary. Experiment 4 used the WBW MT along with the combined vocabulary from the ST and TRT corpora.

Experiments 5 to 8 used SBS machine-translated data. Experiment 5 used no TRT corpus but used the unique words found in TRT_{SBS}, creating the vocabulary $V_{TRT_{SBS}}$. This was done in order to find the impact of including only SBS translated vocabulary. Experiment 6 used TRT_{SBS} as the TRT corpus without adding translated words to the vocabulary. Experiment 7 used the SBS MT along with the combined vocabulary found from the ST and TRT corpora. Experiment 8 used both information from the SBS and WBW MT. Using WBW translated data along with SBS MT can be done since the dictionary used to create the WBW MT was created using the SBS MT.

The ST set size varied from 100 to 1500 sentences for all the experiments. In the following text ST^n corresponds to a subset of the ST set where n is the number of sentences used. Experiments with no ST set included, ST^0, was also performed on Experiment 4, Experiment 7, and Experiments 8. All LMs were built using 3-grams with Kneser-Ney smoothing. The WER experiments were performed three times with different, randomly chosen sentences, creating each ST and SD set, in order to increase the accuracy of the results. An average WER was calculated over the three experiments. This increases accuracy when comparing different experiments especially when the ST set is very sparse. The vocabulary changed for each ST and SD set and the values for words and unique words in Table 1 reflect only one of the three cases. The words and vocabulary sizes for the other two cases were very similar to the one reported in Table 1. Perplexity and out-of-vocabulary (OOV) results reported in this paper also correspond only to the case with ST and SD sets found in Table 1. Each experiment had the interpolation weights optimized on the SD corpus.

The speech recognition experiments were performed using Julius [12] version "rev.3.3p3 (fast)."

5. RESULTS

The WER results from Experiment 1, Experiment 2, Experiment 3, and Experiment 4 are shown in Figure 2. When no manual ST sentences are present and only WBW machine-translated data is used, Experiment 4 gives WER of 67.6%. When 100 ST sentences are used in Experiment 1, the WER *baseline* is 49.6%. Experiment 4 reduces the WER to 46.6% when adding the same number of ST sentences. As more ST sentences are added, the improvement in Experiment 4 reduces and converges with the *baseline* when 500 ST sentences are added to the system. Experiment 2 and Experiment 3 give a small improvement over the *baseline*

Table 9: Perplexity results.

	ST^n				
Experiment no.	ST^0	ST^{100}	ST^{500}	ST^{1000}	ST^{1500}
Experiment 1	NA	30.7	26.4	26.3	26.5
Experiment 3	NA	29.4	26.0	26.1	26.3
Experiment 6	NA	26.6	25.3	25.3	25.4
Experiment 2	NA	58.2	34.2	31.9	30.8
Experiment 4	664.6	50.2	32.6	30.7	29.9
Experiment 5	NA	88.9	43.5	37.7	35.3
Experiment 7	287.0	61.1	38.4	34.1	32.5
Experiment 8	274.8	61.6	38.5	34.4	32.6

Table 10: *OOV* rate (%) with corresponding vocabulary sizes inside parentheses.

	ST^n				
Vocabulary	ST^0	ST^{100}	ST^{500}	ST^{1000}	ST^{1500}
V_{ST^n}	NA (0)	14.0 (190)	6.8 (451)	5.5 (614)	4.6 (805)
$V_{ST^n} + V_{TRT_{WBW}}$	26.8 (3396)	8.4 (3501)	4.8 (3638)	4.0 (3755)	3.4 (3911)
$V_{ST^n} + V_{TRT_{SBS}}$	9.2 (7312)	4.4 (7353)	2.6 (7432)	2.5 (7500)	2.2 (7597)
$V_{ST^n} + V_{TRT_{WBW}} + V_{TRT_{SBS}}$	9.0 (8432)	4.4 (8470)	2.6 (8546)	2.4 (8613)	2.2 (8707)

when the *ST* set is small but converges quickly as more *ST* sentences are added.

The WER results from Experiment 5, Experiment 6, Experiment 7, and Experiment 8 along with the *baseline* in Experiment 1, are shown in Figure 3. When no *ST* sentences are present and only SBS or SBS and WBW machine-translated data is used, Experiment 7 and Experiment 8 give WER of 56.5% and 56.8%, respectively. When 100 *ST* sentences are added to the system and interpolated with the *TRT* corpus in Experiment 7, the WER is 41.9%. Experiment 8 gives a 42.0% WER when 100 *ST* sentences are added to the system. As more *ST* sentences are added, the relative improvement reduces. When 1500 *ST* sentences are used, the WER in Experiment 7 gives 32.5% compared to 32.7% when the *baseline* is used. When the translated vocabulary is alone added, Experiment 5 does not give any significant improvement over the *baseline*. When the vocabulary is fixed to the *ST* set and TRT_{SBS} is used as the *TRT* set, Experiment 6 gives a small improvement over the *baseline*. When *ST* composes of 1500 sentences, the interpolation in Experiment 6 gives a WER of 32.6%. Each experiment was performed three times with different *ST* and *SD* set, and the average WER calculated, as explained before. For example, Experiment 7 shown in Figure 3 gives WER 41.8%, 41.9%, and 42.1%, with an average of 41.9%, when 100 *ST* sentences are used.

When the WER results are more carefully investigated we are able to find out how many more *ST* sentences are needed for Experiment 1 to par Experiment 7. When 100 *ST* sentences are used for Experiment 7 then around 150 *ST* sentences in addition are needed for Experiment 1 to par the WER result of Experiment 7. When 500 *ST* sentences are used for Experiment 7 then around 300 *ST* sentences in addition are needed for Experiment 1 to par the WER results. When 1000 *ST* sentences are used for Experiment 7 then around 200 *ST* sentences in addition are needed for Experiment 1 to par the WER results in Experiment 7.

Perplexity and *OOV* results are shown in Tables 9 and 10, respectively, for some *ST* values. The perplexity results for Experiment 1, Experiment 3, and Experiment 6 should be compared together since the vocabulary is the same for those experiments, V_{ST}. Experiment 2 and Experiment 4 have the same vocabulary, V_{ST} combined with $V_{TRT_{WBW}}$ and should be compared together. For the same reason Experiment 5 and Experiment 7 should be compared together having the same vocabulary, V_{ST} combined with $V_{TRT_{SBS}}$. As shown in Table 9, all perplexity results get improved when a *TRT* corpus is introduced and interpolated with the corresponding *ST* set. The *OOV* rate shown in Table 10 is reduced by adding the unique words found in the *TRT* set to V_{ST} as expected. When the system corresponds to 100 *ST* sentences, the *OOV* rate is reduced from 14.0% to either 8.4% or 4.4% using WBW or SBS MT, respectively. Not applicable (NA) are displayed in Tables 9 and 10 for experiments that have no *ST* sentences and are based solely on the V_{ST} vocabulary and/or are not using any *TRT* corpus, and therefore do not have data to carry out the experiment.

6. DISCUSSION

The improvement of the Icelandic LM with translated English text/data was confirmed by reduction in WER by using either WBW or SBS MT. Experiment 1 should be compared with the other experiments since Experiment 1 does not assume any foreign translation. When the *baseline* in Experiment 1 is compared with the interpolated results using WBW MT in Experiment 4, we get a WER 49.6% reduced to 46.6% respectfully, a 6.0% relative improvement when using 100 ST sentences. The relative improvement reduces as more ST sentences are added to the system and converges to the *baseline* when 500 ST sentences are added to the system. Neither Experiment 2 nor Experiment 3 gives any significant improvement over the *baseline*. This along with the results in Experiment 4 suggests that when WBW translated data is available, both the translated corpus and its vocabulary should be added to the system when the ST sentences are sparse.

The reason why Experiment 8 is not outperforming Experiment 7 is most likely because Experiment 8 is using unique words found in the TRT_{WBW} corpus in addition to the unique words found in Experiment 7. As Table 10 shows, around 1100 new words are added to the vocabulary in Experiment 8 compared to Experiment 7 for all ST set conditions without reducing the OOV rate significantly. Therefore the perplexity rate increases making the speech recognition process more difficult. The unique words found in TRT_{WBW} are therefore not contributing toward better results if vocabulary from TRT_{SBS} is used.

When the *baseline* is compared with the interpolated results using SBS MT in Experiment 7, we get a WER 49.6% reduced to 41.9% respectfully, a 15.5% relative improvement when 100 ST sentences are added to the system. Improvements by merging the vocabulary from the TRT_{SBS} and V_{ST} is confirmed by comparing Experiment 6 and Experiment 7 for all ST sets. The WER improvement of the SBS MT over the WBW MT is confirmed for all the ST sets as the BLEU evaluation results in Section 3.1 suggests. This can be seen by comparing Experiment 4 in Figure 2 with Experiment 7 in Figure 3. The improvement is as well confirmed with perplexity results when Experiment 3 and Experiment 6 are compared in Table 9. When the vocabulary is kept the same as in the case of Experiment 1, Experiment 3, and Experiment 6 the proposed methods always outperform the baseline perplexity results.

7. CONCLUSIONS

The results presented in this paper show that an LM can be improved considerably using either WBW or SBS translation. This especially applies when developing a prototype system where the amount of target domain sentences is very limited. The effectiveness of the WBW and SBS translation methods was confirmed for English to Icelandic for a weather information task. The convergence point of these methods with the baseline was around 400 and 1500 manually collected sentences for the WBW and the SBS translation methods respectfully. In order to get significant improvement, a good (high BLEU score) MT system is needed. The WBW translation is especially important for resource-deficient languages that do not have SBS machine translation tools available. It is believed that a high BLEU score can be obtained with WBW MT for very closely related language pairs and between dialects. Confirming the effectiveness of the WBW and the SBS translation methods for other language pairs is left as future work, as is applying the rule based WBW and SBS translation methods to a larger domain, for example broadcast news. Future work also involves an investigation of other maximum a posteriori adaptation methods such as [13] and methods like the ones described in [14–16] that selects a relevant subset from a large text collection such as the World Wide Web to aid sparse target domain. These methods assume that a large text collection is available in the target language but we would like to apply these methods to extract sentences from the TRT corpus. Since the acoustic model is only built from 3.8 hours of acoustic data which gives rather poor results we would like to either collect more Icelandic acoustic data or use data from foreign languages to aid current acoustic modeling.

ACKNOWLEDGMENTS

The authors would like to thank Dr. J. Glass and Dr. T. Hazen at MIT and all the others who have worked on developing the Jupiter system. They also would like to thank Dr. Edward W. D. Whittaker for his valuable input. Special thanks to Stefan Briem for his English to Icelandic machine translation tool and allowing to use his machine translation results. This work is supported in part by 21st Century COE Large-Scale Knowledge Resources Program.

REFERENCES

[1] M. Adda-Decker, "Towards multilingual interoperability in automatic speech recognition," *Speech Communication*, vol. 35, no. 1-2, pp. 5–20, 2001.

[2] T. Schultz and A. Waibel, "Language-independent and language-adaptive acoustic modeling for speech recognition," *Speech Communication*, vol. 35, no. 1-2, pp. 31–51, 2001.

[3] S. Khudanpur and W. Kim, "Using cross-language cues for story-specific language modeling," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP '02)*, vol. 1, pp. 513–516, Denver, Colo, USA, September 2002.

[4] W. Kim and S. Khudanpur, "Cross-lingual latent semantic analysis for language modeling," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 1, pp. 257–260, Montreal, Canada, May 2004.

[5] H. Nakajima, H. Yamamoto, and T. Watanabe, "Language model adaptation with additional text generated by machine translation," in *Proceedings of the 19th International Conference on Computational Linguistics (COLING '02)*, vol. 2, pp. 716–722, Taipei, Taiwan, August 2002.

[6] M. Paulik, S. Stüker, C. Fügen, T. Schultz, T. Schaaf, and A. Waibel, "Speech translation enhanced automatic speech recognition," in *Proceedings of IEEE Workshop on Automatic*

Speech Recognition and Understanding (ASRU '05), pp. 121–126, San Juan, Puerto Rico, November-December 2005.

[7] V. Zue, S. Seneff, J. R. Glass, et al., "JUPITER: a telephone-based conversational interface for weather information," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 1, pp. 85–96, 2000.

[8] S. Briem, "Machine Translation Tool for Automatic Translation from English to Icelandic," Iceland, 2007, http://www.simnet.is/stbr/.

[9] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: a method for automatic evaluation of machine translation," in *Proceedings of the 40th Annual Conference of the Association for Computational Linguistics (ACL '02)*, pp. 311–318, Philadelphia, Pa, USA, July 2002.

[10] E. Rögnvaldsson, *Islensk hljodfraedi*, Malvisindastofnun Haskola Islands, Reykjavik, Iceland, 1989.

[11] S. Young, G. Evermann, T. Hain, et al., "The HTK Book (Version 3.2.1)," 2002.

[12] A. Lee, T. Kawahara, and K. Shikano, "Julius—an open source real-time large vocabulary recognition engine," in *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH '01)*, pp. 1691–1694, Aalborg, Denmark, September 2001.

[13] M. Bacchiani and B. Roark, "Unsupervised language model adaptation," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 1, pp. 224–227, Hong Kong, April 2003.

[14] R. Sarikaya, A. Gravano, and Y. Gao, "Rapid language model development using external resources for new spoken dialog domains," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 1, pp. 573–576, Philadelphia, Pa, USA, March 2005.

[15] A. Sethy, P. Georgiou, and S. Narayanan, "Selecting relevant text subsets from web-data for building topic specific language models," in *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL '06)*, pp. 145–148, New York, NY, USA, June 2006.

[16] D. Klakow, "Selecting articles from the language model training corpus," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*, vol. 3, pp. 1695–1698, Istanbul, Turkey, June 2000.

EURASIP Journal on Image and Video Processing

Special Issue on Image and Video Processing for Cultural Heritage

Call for Papers

Digital imaging and 3D modeling are nowadays extensively employed to capture, conserve, describe, and render cultural artifacts such as buildings and monuments, archaeological sites, artworks, manuscripts, books, and other objects of artistic, historical or archaeological interest. Computer vision, graphics, image and signal processing are essential instruments for virtual and physical restoration, analysis, and documentation of the artifact content. The ultimate goal is to facilitate the access and study of our cultural heritage by the public and scholars alike and to ensure its preservation for the future. Visual processing of cultural heritage data does not merely exploit and apply standard techniques, but often entails original research specific to this domain.

The creation of digital libraries, delivering to the users' content that fits their needs, requires the development of powerful indexing and retrieval tools and creates a need for protection against improper usage and preservation of integrity and authenticity. The often low quality and high complexity of the content requires multimodal acquisition to enrich documentation, recover masked information, and facilitate analysis. This special issue aims to address these challenging issues. High-quality, original contributions on the following (non exhaustive) list of topics are solicited:

- High resolution 2D and 3D digital representations, correction of degradations and quality evaluation
- Multimodal, multiresolution, and HDR imaging; data registration and fusion
- Signal, image processing, and 3D modeling to assist physical restoration
- Extraction, recognition, classification, and enhancement of features, structures, and contents
- Digital restoration of damaged artworks (films, photographs, paintings, frescos, manuscripts, etc.)
- Storage, handling, transmission, processing, and visualization of large datasets
- Visualization of archaeological sites: temporal evolution, uncertainty in the model, GIS layers
- Large-scale multimedia databases of artworks; archival, indexing, and retrieval; copyright and IPR management
- Automatic artist/creator or artistic style recognition, detection of forgery/fakes, and dating of artwork
- User-centered visual applications for museums, digital art repositories, and edutainment (VR, AR, etc.)

Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://www.hindawi.com/journals/ivp/guidelines.html. Authors should follow the EURASIP Journal on Image and Video Processing manuscript format described at http://www.hindawi.com/journals/ivp/. Prospective authors should submit an electronic copy of their complete manuscripts through the EURASIP Journal on Image and Video Processing manuscript tracking system at http://mts.hindawi.com/, according to the following timetable:

Manuscript Due	February 1, 2009
First Round of Reviews	May 1, 2009
Publication Date	August 1, 2009

Lead Guest Editor

Vincent Charvillat, Département Informatique et Mathématiques Appliquées, l'Enseeiht Institut de Recherche en Informatique de Toulouse (IRIT), UMR 5505-CNRS, 31071 Toulouse cedex 4, France; charvi@enseeiht.fr

Guest Editors

Anna Tonazzini, Istituto di Scienza e Tecnologie dell'Informazione CNR, 56124 Pisa, Italy; anna.tonazzini@isti.cnr.it

Luc Van Gool, Computer Vision Laboratory, Swiss Federal Institute of Technology Zurich (ETH Zurich), 8092 Zurich, Switzerland; ESAT-PSI/Visics, Katholieke Universiteit Leuven, 3001 Heverlee, Belgium; vangool@vision.ee.ethz.ch

Nikos Nikolaidis, Department of Informatics, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece; nikolaid@aiia.csd.auth.gr