7-2 Speech-to-Speech Translation System Field Experiments in All Over Japan

YASUDA Keiji and MATSUDA Shigeki

We explain field experiments conducted during the 2009 fiscal year in five areas of Japan. We also show the experiments of evaluation and data selection method from speech translation field data. The data selection method selects useful data from filed data by using a development data set. According to the experimental results, the proposed data selection method gives the improvement of the speech-to-speech translation systems.

Keywords
Speech translation field experiment, Statistical machine translation, Speech translation system, Speech translation field data

1 Speech translation field experiments

In this paper we first explain the speech translation field experiments conducted in 2009 [1], and then the method for improving the speech translation system using the data set collected in the field experiments. Aiming at the significant enhancement of the translation accuracy of automatic speech translation technology and the early realization of a service that utilizes this technology for foreign visitors to Japan, the Ministry of Internal Affairs and Communications conducted field experiments, titled "Field Testing of Automatic Speech Translation Technology and Its Contribution to Local Tourism" (total works budget of 985 million yen), contracting them to privately-owned corporations.

The field experiments were performed with about 1,700 terminals in about 370 tourist facilities in five areas, as shown in Fig. 1, for the four languages of Japanese, English, Chinese, and Korean. About 200,000 accesses were recorded in the period of the experiments. This was the world's first large-scale field experiment conducted under practical conditions. The National Institute of Information and Communications Technology provided its speech translation technology to all of the corporations that were contracted by the Ministry, and offered full cooperation for the development and operation of the experiment system and data analysis.

2 System configuration

Figure 2 shows an outline of the configuration of the field experiment system developed by each project area. The speech translation terminal consists of a smart phone and a laptop PC, and 300 to 500 terminals were installed in each area. A speech given to the terminal is sampled at 16 kHz and sent in an ADPCM format to a speech translation server. The speech translation server consists of a speech recognition server, a machine translation server, and a speech synthesis server, which are all available for every language. The translation result is sent to the terminal in the form of text or a synthesized speech. Also, the speech input, result of speech recognition, and translation result are saved, together with the date, terminal ID, and specified language, in a log file in the system.

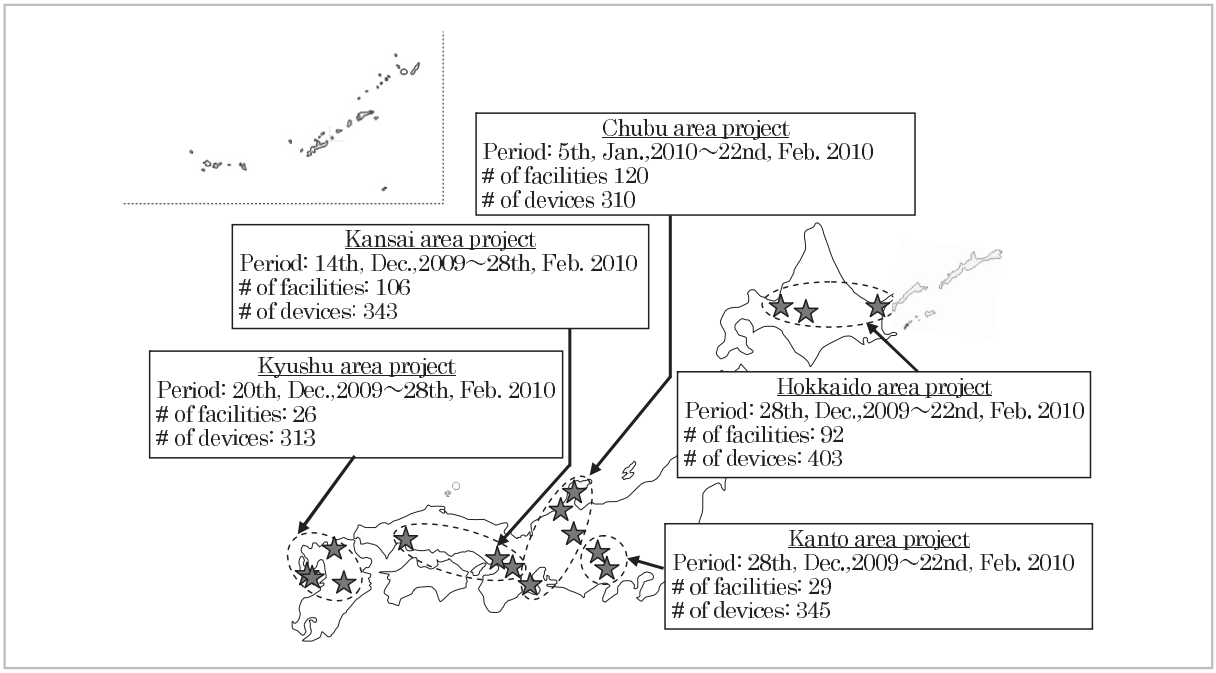

Fig.1 *Outline of field experiments in 5 regions*

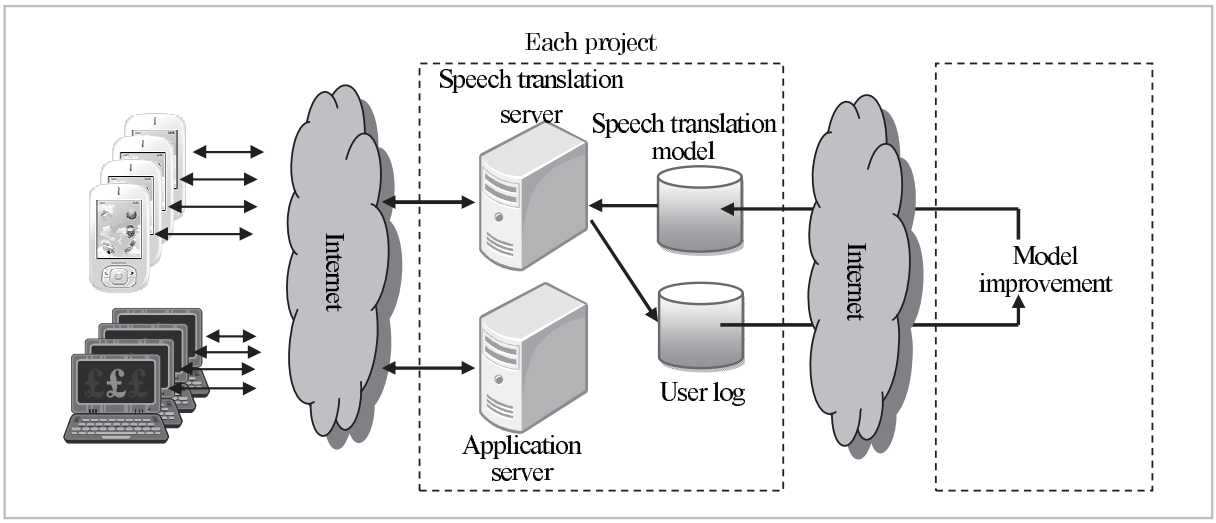

Fig.2 *System configuration of speech translation field experiments*

2.1 Outline of speech recognition system

The speech recognition system used in this project consists of a front end unit and decoder unit. A particle filter in the front end unit successively estimates the noise strength, which changes with time, to suppress non-stationary noise [2]. For the decoder unit, a Hidden Markov Model (HMM) is used as an acoustic model, and a multi-class composite N-gram [3], an extended version of word class N-gram, is used as the language model. These are used for two-path speech recognition. In the first path, the acoustic model and 2-gram language model are used to create a word graph. In the second path, a trigram language model is used for word lattice rescoring and to look for a recognition result.

The acoustic model is trained by using a 400-hour speech corpus of about 4,500 adult and elder persons. It was expected that the speech recognition system was to be used in a

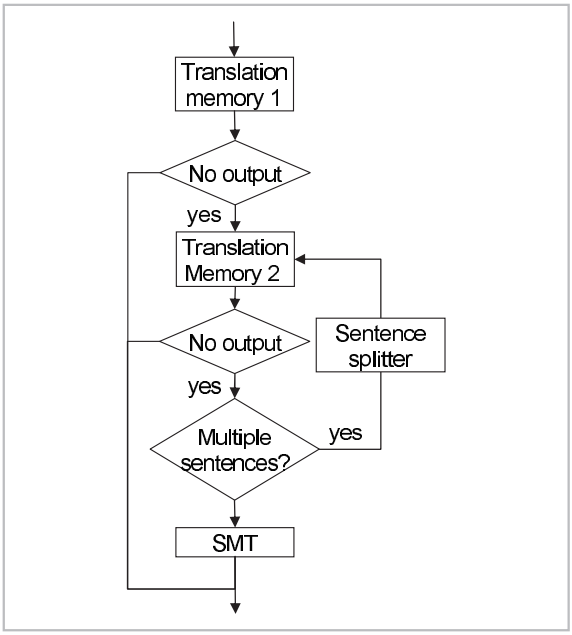

Fig.3 *Process flow of machine translations*

noisy environment, e.g. outdoors, for the field experiments. We therefore estimated the acoustic model using training data in which automobile noise and other noises recorded in various places, such as streets and railway stations, were randomly superimposed at an S/N ratio of 10–30 dB.

We then estimated multi-class composite 2-gram and multi-class 3-gram language models using a text corpus of about 740,000 sentences collected from travel conversations. The lexicon size was about 50,000 words.

2.1.1 Customization of language model for each area

We customized the language model for each area, using proper nouns and regional expressions (dialects) specific to the area. The proper nouns (about 5,000 words) used in these areas were categorized by place name, facility name, and others and the language probability of typical words in the basic dictionary was assigned to each category. Also the regional expressions (about 3,000 sentences) were used for the training of the word N-gram and were combined linearly with the basic language model for the adaptation of the model.

2.2 Overview of machine translation system

Figure 3 shows details of the processing of the machine translation unit in the speech translation system that consists of speech recognition, machine translation and speech synthesis units. The machine translation unit consists mostly of a statistical machine translation unit and two translation memories. For the statistical machine translation system, a phrase-based statistical machine translation framework [4] was used. In this framework, the probability of a word sequence (e) translated in a target language from its source word sequence (f) is calculated from the following equation.

$$p(e|f) = \frac{\exp\left(\sum_{i=1}^{M} \lambda_i h_i(e,f)\right)}{\sum_{e'} \exp\left(\sum_{i=1}^{M} \lambda_i h_i(e',f)\right)} \qquad (1)$$

Here, e' represents a candidate translation of f. $h_i\ (e,f)$ is a feature function obtained from a learning corpus. There are eight feature functions, including translation probabilities (translation models) of a word or phrase from target language to source language or from source language to target language, and language models of target languages [5]. λ_i and M are a weight of each feature function and the number of feature functions (8), respectively.

Assuming that the denominator of Equation (1) is constant, we derive a translation result ê from the following Equation (2).

$$\hat{e}(f, \lambda_1^M) = \mathrm{argmax}_e \sum_{i=1}^{M} \lambda_i h_i(e,f) \qquad (2)$$

We used the MOSES tool kit [5] and SRILM tool kit [6] for the learning of the translation models and language models.

In the field experiments, the following two kinds of data were collected prior to the experiments in each area.

-Proper nouns: Several thousand regional nouns. Their translations (in English, Chinese and Korean) and their categories were also collected.

-Regional expressions: Several thousand regional expressions (texts), collected through

prior interviews in the places where each terminal was to be installed. Their translations (in English, Chinese and Korean) were also collected. The data also included expressions that are necessary for the business of stores where terminals were installed.

The proper nouns are used in the translations by replacing them with tokens of the corresponding category in advance as proposed by the reference [7]. In the learning of the models, not only the BTEC corpus [8] but also regional expressions are used. First, the above-mentioned tool kits are used for the learning of the feature function for each of the BTEC corpus and regional expressions, and the two feature functions are linearly combined:

$$\begin{aligned}\mathbf{h}_{baseline}(e,f) \\ = \quad & \mu\mathbf{h}_{btec}(e,f) \\ & +(1-\mu)\mathbf{h}_{regional}(e,f)\end{aligned} \tag{3}$$

Here, $\mathbf{h}_{btec}$ is the eight feature functions of Equation (1) obtained from the BTEC corpus, and $\mathbf{h}_{regional}$ is the eight feature functions obtained from the regional expressions. $\mathbf{h}_{baseline}$ is the eight feature functions obtained by the linear combination of these two. For the weight μ, we will explain in Subsection **4.1**.

The regional expressions can be used in another way. Namely, they are incorporated into the corpus, which is then used for the learning of a model. However if there are additional regional expressions, the model learning has to be made again with all the data. On the other hand, with the method of Equation (3), one can make the model learning only with the small-size data of the regional expressions, which provides for easy maintenance.

The regional expressions are also used in the translation memories. Translation memory 1 in Fig. 3 uses the BTEC corpus and translation memory 2 uses the regional expressions collected from each area.

3 Use of data collected in field experiment

In our proposed method, we use the results of back translations from translated texts to the source language in order to choose the data for model adaptation. In Subsection **3.1**, we explain supervised filtering which is based on the proposed method or two conventional methods. In Subsection **3.2** we describe a supervised filtering method, which is based on conventional methods. In Subsection **3.3**, we explain how the selected data is used for model adaptation.

3.1 Unsupervised filtering

3.1.1 Filtering by normalized translation score

In a conventional method [9] that uses normalized translation scores, the following formula is used to calculate the normalized translation score (S_{trans}) of data, and if it is larger than a threshold, the data is used for model adaptation.

$$S_{trans} = p(e|f)^{\frac{1}{n_e}} \tag{4}$$

Here, n_e represents the number of words in the translation results and $p(e|f)$ shows the translation probability calculated by Equation (1).

3.1.2 Filtering by source language perplexity

In another conventional method [10] that uses source language perplexity, the following formula is used to calculate the perplexity ($S_{pp}(f)$) of the data, and if it is larger than a certain threshold, the data is used for model adaptation.

$$S_{pp} = p(f)^{-\frac{1}{n_f}} \tag{5}$$

Here, n_f represents the number of words in the input sentence and $p(f)$ shows the perplexity of the sentence given by the language model of the source language. In this paper, we use the language model for which the learning is made with the source language side of the BTEC corpus in Section **2**.

3.1.3 Proposed method

In our proposed method, the result of a automatically translated text from the source language to the target language is automatically back-translated to the source language. Then, using the result (f) of the speech recognition

of the source as a reference translation, an automatic evaluation score for the result (f') of the back-translation is calculated. We use the following PER (position independent word error rate) as the automatic translation evaluation value:

$$S_{src\_per} = PER(f, f') \tag{6}$$

Here, PER is the word error rate calculated from an evaluation of the speech recognition ignoring the word order. For the evaluation of machine translations, the BLEU [11] score is often used as automatic evaluation score. However if BLEU, which should be applied to speech, is applied to the present cases, the score is 0 in most cases and we cannot rate the translation. This is a problem of mismatching between the translation quality and the evaluation method, and we employ the PER which has a relatively higher compatibility between them.

3.2 Supervised filtering

Supervised filtering is a method of selecting high-quality translation through the automatic translation evaluation of each sentence in the results of machine translations. For the evaluation, manually-made reference translations are used. Since the aim of the unsupervised adaptation that we use in this paper is adaptation without using manually-made reference translations, the use of these reference translations for filtering is not realistic. In this paper, however, we also conducted supervised filtering to perform additional experiments to the previous study [12] and to compare the performance of the supervised filtering with that of the unsupervised filtering.

In the previous study [12], BLEU [11] was used for automatic translation evaluation. In this paper, on the other hand, *PER* (S_{tgt_per}) calculated from the following formula is used.

$$S_{tgt\_per} = PER(e, e') \tag{7}$$

3.3 Adaptation method

The selected data is used in the following manner, together with the regional expression data mentioned in Section **2**.

Step 1 The obtained data and the regional expressions mentioned in Section **2** are combined into an adaptation corpus.

Step 2 The corpus obtained in Step 1 is used for the learning of the eight feature functions ($\mathbf{h}_{field}(e, f)$) of Equation (1).

Step 3 The model ($\mathbf{h}_{btec}(e, f)$) for which the learning is made with the BTEC corpus is linearly combined with $\mathbf{h}_{field}(e, f)$ to obtain an adaptation model ($\mathbf{h}_{adapted}(e, f)$).

$$\begin{aligned}&\mathbf{h}_{adapted}(e, f)\\ &\quad = \mu\mathbf{h}_{btec}(e, f) + (1 - \mu)\mathbf{h}_{field}(e, f)\end{aligned} \tag{8}$$

The value of the weight μ will be discussed in the Subsection **4.1**.

4 Experiments

Next we examine how to improve the performance of the speech translation system by using actual data collected in the field experiments. In ordinary studies on speech translation systems, the data is manually annotated (transcribed and translated) and the arranged data is used for the re-learning of the system. This method is very effective but requires a great deal of cost and time for the data annotation. To circumvent this problem, we considered using the data without manual annotation.

4.1 Experiment conditions

In the preliminary experiment, the subjective evaluation of the translation quality was worst in Hokkaido and highest in Kyushu. We hence used the data obtained in Hokkaido and Kyushu. We focused on translations from Japanese to English since the needs for translation in this direction was high in every area and since we could collect the largest volume of data for this translation direction.

The Japanese-English BTEC corpus used for model learning contains 691,829 sentences, and the regional expressions obtained in Hokkaido and those in Kyushu were 3,000 sentences and 5,095 sentences respectively.

For the evaluation of the machine translations, we used 100 sentences randomly picked

up from each region's data. The perplexity of the test set of Hokkaido in the language model for which the learning is performed using only the Japanese BTEC corpus was 40.98 and that of Kyushu was 19.36. On the other hand, the perplexity of the test set of Hokkaido with the baseline language model obtained by Equation (3) was 42.64 and that of Kyushu was 17.17. We thus see that the test set of Hokkaido was more difficult than that of Kyushu.

In the preliminary experiments and the present experiments, we made subjective evaluations of the translation quality using five grades: S (Perfect), A (Correct), B (Fair), C (Acceptable), and D (Nonsense). In the evaluation shown in Subsection **4.2**, we employed text input without using speech recognition for the machine translation of a test sentence.

The values of the eight weights λ_i of the feature functions in Equation (2) were determined by MERT (Minimum Error Rate Training) [13] with a BTEC development set of 500 sentences and were used for all the experiments. In the same manner, the weight μ in Equations (3) and (8) was set to 0.9 according to the preliminary experiment using the BTEC development set. These values were the same as those used in the field experiments. The reasons that we chose these values in the field experiments were as follows.

-Development set of actual data cannot be sufficiently obtained before the field experiments.

-The speech translation system is used differently depending on when it is used. For example, the system was used during the Snow Festival. Therefore, the parameter tuning which is based on a small-size development set collected in a certain time period may cause over-adaptation.

-The parameter tuning based on the BTEC development set can maintain at least the performance of the speech domain of the BTEC.

For optimal parameter setting and field data obtained in the entire period of the field experiments should be used for random sampling of development sets from the data, and the weights λ_i and μ need to be tuned individually by the development sets. In this experiment, however, we used the same setting as that used in the field experiments since we needed perform experiments under a practical situation where our proposed method will be used.

4.2 Experiment results

Table 1 shows the evaluation results of the case where the field data obtained was all used

Table 1 *Evaluation results of supervised/unsupervised adaptation (with no data filtering)*

Project Area	System Type	Additional Field Data			Ratio (%)			
		Transcription	Translation	Size (# of sentences)	S	S, A	S, A, B	S, A, B, C
Hokkaido	Baseline	N/A	N/A	0	29	38	55	62
	Baseline + unannotated data	ASR	MT	9602	29	38	53	61
	Baseline + unannotated data 1	Manual	MT	10009	31	39	51	62
	Baseline + unannotated data 2 (Upper bound)	Manual	Manual	10335	34	44	61	68
Kyushu	Baseline	N/A	N/A	0	50	62	72	76
	Baseline + unannotated data	ASR	MT	9722	50	60	71	76
	Baseline + unannotated data 1	Manual	MT	10337	49	62	72	77
	BL + unannotated data 2 (Upper bound)	Manual	Manual	14138	55	64	74	79
	Baseline with 3000 regional expressions	N/A	N/A	0	47.7	60.0	69.0	73.3

without data filtering. For each area, the first line of the table shows the baseline and the second is the result of unsupervised adaptation using all the usable field data. The third line shows the results of the case where the transcription of input speech was made manually and machine translation was utilized to use the target language data in the adaptation, i.e. when partially-supervised adaptation is made. The fourth line presents the results of supervised adaptation, where the transcription and translation were both made manually. This result of supervised adaptation gives the upper bound of the performance improvement.

The data size varies even in the same area depending on the conditions, because data was not used in the adaptation if output could not be obtained in the speech recognition or machine translation.

As mentioned earlier, Kyushu's data contains more regional expressions than Hokkaido's. To find an influence of the number of regional expressions, we show in the last line of Table 1 the results of the case where the number of regional expressions in Kyushu was adjusted by random sampling to 3,000 sentences, the same number as in Hokkaido. We conducted the random sampling three times and averaged the results of the subjective evaluation.

The white cells in the table indicate higher performance than the baseline, the gray cells indicate equivalent performance with the baseline, and the dark gray cells indicate lower performance than the baseline*. From Table 1 we see no improvements in the unsupervised adaptation. The partially-supervised adaptation showed performance improvements in one area but performance degradation in the other. On the other hand, the supervised adaptation had performance improvements in both areas. We see from these results that, if we use the field data without filtering it, there could be performance degradation, in particular in the unsupervised adaptation. The filtering of the data should therefore be employed to prevent this degradation and the effect of our proposed method is shown in Table 2.

We finally compared the baseline data of the two regions. In Kyushu's data, the performance degraded when the number of regional expressions was decreased to 3,000. However it was still much higher than that of Hokkaido's data which contains the same number of regional expressions. We could therefore conclude that Kyushu's test set was easier to translate than Hokkaido's test set. This difference could be attributed to the fact that, in the field experiments in Kyushu, the terminals were placed at airports and tourist spots with guides under relatively controlled conditions.

Table 2 shows the evaluation results obtained by conventional methods and our proposed method, where the speech translation field data is filtered for unsupervised adaptation in the presence of speech recognition errors and machine translation errors. The word error rate of the speech recognition system is 29.9% in Hokkaido and 20.3% in Kyushu.

For the proposed method, we set the filtering threshold to be 0.1, 0.2, or 0.4. To make a fair comparison, the thresholds in the conventional methods were adjusted in such a way that the adjusted threshold would give the same number of sentences as that obtained with the thresholds of 0.1, 0.2, or 0.4 in the proposed method.

Let us first compare the data quantities when the threshold is set to 0.1 in the proposed method. The filtering of Hokkaido's data gives 1,244 sentences (12.7% of entire data obtained in Hokkaido), while the filtering of Kyushu's data gives 4,560 sentences (46.9% of entire data obtained in Kyushu.) This difference was caused by the following:

-In comparison to Hokkaido, Kyushu's data had a low word error rate and hence the speech recognition errors had little influence on the translation quality. As a result, the back-translated texts and the input texts were in relatively good agreement.

-In comparison to Hokkaido, Kyushu's

* The same colors are used in Tables 2 and 3.

data contained a lot of easy-to-translate data and was hence translated relatively correctly from source to target language as well as from target to source language. As a result, the back translated texts and the input texts were in relatively good agreement.

Next we focus on the data set of Hokkaido. With the proposed method, the system performance improved from the baseline in almost all cases except the case of the threshold equal to 0.1, where the performance degraded. The performance improved also when the normalized translation score (S_{trans}) or source language perplexity (S_{pp}) was used, although the improvement from the proposed method was greater. Comparing the proposed method (S_{src_per}) to the supervised filtering (S_{tgt_per}), we see that the proposed method improved performance even though it was unsupervised, more than the supervised filtering method. With the condition $S_{src\_per} <= 0.2$, the total ratio of S, A, B, and C was higher than the corresponding Upper bound shown in Table 1. This indicated that when there is no translation error or speech recognition error the translation quality could be improved by eliminating the training sentences that are not included in the domain [14]. This could accidentally happen in the case of $S_{src\_per} <= 0.2$. We considered that even the Upper bound in Table 1 could be exceeded

Table 2 *Evaluation results of unsupervised adaptation (with data filtering)*

Project Area	Additional field data		Ratio (%)			
	Filtering function	Size (# of sentences)	S	S, A	S, A, B	S, A, B, C
Hokkaido	N/A (Baseline)	0	29	38	55	62
	S_{trans} (eq. 4)	1244	32	41	54	64
		1861	30	40	52	61
		3565	32	42	53	63
	S_{pp} (eq. 5)	1244	30	41	53	65
		1861	30	41	53	65
		3565	30	40	53	62
	$S_{src\_per} <= 0.1$ (eq. 6)	1244	31	40	55	66
	$S_{src\_per} <= 0.2$ (eq. 6)	1861	32	41	56	69
	$S_{src\_per} <= 0.4$ (eq. 6)	3565	32	41	56	66
	S_{tgt_per} (eq. 7)	1244	32	40	54	64
		1861	30	40	53	63
		3565	31	41	53	63
Kyushu	N/A (Baseline)	0	50	62	72	76
	S_{trans} (eq. 4)	4560	49	62	72	75
		5274	49	62	72	75
		6699	49	61	71	77
	S_{pp} (eq. 5)	4560	48	60	70	74
		5274	48	59	69	74
		6699	48	59	71	74
	$S_{src\_per} <= 0.1$ (eq. 6)	4560	49	60	70	74
	$S_{src\_per} <= 0.2$ (eq. 6)	5274	49	61	71	75
	$S_{src\_per} <= 0.4$ (eq. 6)	6699	51	61	71	74
	S_{tgt_per} (eq. 7)	4560	50	62	72	76
		5274	50	61	71	75
		6699	49	60	71	75

by using the method of [14] for data selection.

4.2.1 Detailed analysis of Kyushu's data set

Table 3 shows the results of the proposed method where the field data was used only for the language model adaptation and/or the translation model adaptation. We see from the table that, in Hokkaido's data set, the performance was most improved when the adaptation was made for both the language and translation models and the improvement was small when the adaptation was conducted only for the language model or for the translation model.

In Kyushu's data set, on the other hand, the adaptation of only the language model degraded performance, while the adaptation of only the translation model improved performance. In actual operations, using a development data set to determine a model which is suitable for adaptation would prevent performance degradation and improve performance.

5 Summary

We proposed a machine translation adaptation method using field data from speech translations, which contains speech recognition data and machine translation data. In our proposed method, the results of the machine translation were translated back to the source language, and the back-translated text was compared with the results of the speech recognition. The data was employed for adaptation if its input text (results of the speech recognition) and the back-translated text are close to each other.

We conducted experiments using the data

Table 3 *Evaluation results of each model adaptation (with data filtering)*

Project Area	Filtering function	Additional field data		Ratio (%)			
		Used for LM training	Used for TM training	S	S, A	S, A, B	S, A, B, C
Hokkaido	N/A (Baseline)	No	No	29	38	55	62
	$S_{src\_per} <= 0.1$	Yes	Yes	31	40	55	66
	$S_{src\_per} <= 0.2$	Yes	Yes	32	41	56	69
	$S_{src\_per} <= 0.4$	Yes	Yes	32	41	56	66
	$S_{src\_per} <= 0.1$	Yes	No	32	40	54	63
	$S_{src\_per} <= 0.2$	Yes	No	32	41	56	64
	$S_{src\_per} <= 0.4$	Yes	No	31	41	55	64
	$S_{src\_per} <= 0.1$	No	Yes	31	40	54	64
	$S_{src\_per} <= 0.2$	No	Yes	32	40	55	64
	$S_{src\_per} <= 0.4$	No	Yes	32	40	54	63
Kyushu	N/A (Baseline)	No	No	50	62	72	76
	$S_{src\_per} <= 0.1$	Yes	Yes	49	60	70	74
	$S_{src\_per} <= 0.2$	Yes	Yes	49	61	71	75
	$S_{src\_per} <= 0.4$	Yes	Yes	51	61	71	74
	$S_{src\_per} <= 0.1$	Yes	No	47	57	68	73
	$S_{src\_per} <= 0.2$	Yes	No	47	57	68	73
	$S_{src\_per} <= 0.4$	Yes	No	47	57	68	73
	$S_{src\_per} <= 0.1$	No	Yes	51	62	73	76
	$S_{src\_per} <= 0.2$	No	Yes	50	61	73	76
	$S_{src\_per} <= 0.4$	No	Yes	52	63	73	76

that was obtained in the speech translation field experiments conducted in 2009. As a result we found that the proposed method improved the translation performance for the data obtained in the Hokkaido area whose baseline performance was low. On the other hand the baseline performance for the data obtained in Kyushu was relatively high and the translation performance was quite degraded by the present method. However if we made the adaptation not for the language model, but only for the translation model, the performance was improved to a certain degree.

In actual operations, it is expected that system performance would be improved by the proposed method without the manual transcription or manual creation of parallel texts, although we need to prepare a development set to determine the threshold for data filtering and to determine a model to which the adaptation can be made.

The result of a questionnaire survey of the field experiments shows that the performance of current speech translation systems is not high enough to satisfy everyone. The present method has a function of adjusting the probability of each model using field data but cannot handle unknown words contained in the field data. For further improvements of the system, we need to incorporate a mechanism for acquiring proper nouns automatically from the Web into the speech translation system and to continue to collect field data.

References

1 H. Kawai, R. Isitani, K. Yasuda, E. Sumita, M. Uchiyama, S. Matsuda, Y. Ashikari, and S. Nakamura, "An overview of a nation-wide field experiment of speech-to-speech translation in fiscal year 2009," Proc. of 2010 Autumn Meeting of the ASJ, pp.99–102, 2010.

2 M. Fujimoto and S. Nakamura, "A non-stationary noise suppression method based on particle filtering and polyak averaging," The IEICE Transactions on Information and Systems, Vol. E89-D, No. 3, pp. 922–930, 2006.

3 H. Yamamoto and Y. Sagisaka, "Multi-Class Composite N-gram Language Model Based on Connection Direction," The IEICE Transactions on Information and Systems, Vol.J83-D-II, No.11, pp.2146–2151, 2000.

4 P. Koehn, F. J. Och, and D. Marcu, "Statistical Phrase-Based Translation," Proc. of Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL), pp. 127–133, 2003.

5 P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, "Moses: Open source toolkit for statistical machine translation," Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, pp. 177–180, Association for Computational Linguistics, June 2007.

6 A. Stolcke, "SRILM-an extensible language modeling toolkit," Proceedings of the International Conference on Spoken Language Processing, pp. 901–904, 2002.

7 H. Okuma, H. Yamamoto, and E. Sumita, "Introducint a translation dictionary into phrasebased smt," The IEICE Transactions on Information and Systems, vol. 91-D, no. 7, pp. 2051–2057, 2008.

8 G. Kikui, S. Yamamoto, T. Takezawa, and E. Sumita, "Comparative study on corpora for speech translation," IEEE Transactions on Audio, Speech and Language Processing, vol. 14(5), pp. 1674–1682, 2006.

9 N. Ueffing, G. Haffari, and A. Sarkar, "Semi-supervised model adaptation for statistical machine translation," Machine Translation, Vol. 21, No. 2, pp. 77–94, 2007.

10 K. Yasuda, R. Zhang, H. Yamamoto, and E. Sumita, "Method of selecting training data to build a compact

and efficient translation model," Proceedings of the Third International Joint Conference on Natural Language Processing, pp. 655–660, 2008.

11 K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 311–318, 2002.

12 N. Bach, R. Hsiao, M. Eck, P. Charoenpornsawat, S. Vogel, T. Schultz, I. Lane, A. Waibel, and A. W. Black, "Incremental adaptation of speech-to-speech translation," Proceedings of NAACL HLT 2009, pp. 149–152, 2009.

13 F. Och, "Minimum error rate training in statistical machine translation," Proc. of 41st Annual Meeting of the Association for Computational Linguistics (ACL), pp. 160–167, 2003.

14 K. Yasuda, H. Yamamoto, and E. Sumita, "Training set selection for building compact and efficient language models," The IEICE Transactions on Information and Systems, Vol. 92-D, No. 3, pp. 506–511, 2009.

(Accepted June 14, 2012)

***YASUDA Keiji**, Dr. Eng.*

Senior Researcher, Multilingual Translation Laboratory, Universal Communication Research Institute

Machine Translation, Natural Language Processing

***MATSUDA Shigeki**, Ph.D.*

Senior Researcher, Spoken Language Communication Laboratory, Universal Communication Research Institute

Signal Processing, Speech Recognition