

Providing Medical Information in Braille: Research and Development of Automatic Braille Translation Program for Japanese “eBraille”

Aki Sugano, Mika Ohta, Mineko Ikegami, Kenji Miura, Sayo Tsukamoto, Akihiro Ichinose,
Toshiko Ohshima, Eiichi Maeda, Masako Matsuura and Yutaka Takaoka

Abstract—Along with the advances in medicine, providing medical information to individual patient is becoming more important. In Japan such information via Braille is hardly provided to blind and partially sighted people. Thus we are researching and developing a Web-based automatic translation program “eBraille” to translate Japanese text into Japanese Braille. First we analyzed the Japanese transcription rules to implement them on our program. We then added medical words to the dictionary of the program to improve its translation accuracy for medical text. Finally we examined the efficacy of statistical learning models (SLMs) for further increase of word segmentation accuracy in braille translation. As a result, eBraille had the highest translation accuracy in the comparison with other translation programs, improved the accuracy for medical text and is utilized to make hospital brochures in braille for outpatients and inpatients.

Keywords—Automatic Braille translation, Medical text, Partially sighted people.

I. INTRODUCTION

RECENT advances in medical care, developed as a result of input from patients, have encouraged the use of written materials to explain tests or treatments to patients. Thus, patients receive many papers and documents: informed consent; clinical path, which provide explanations of treatments; plans or instructions for hospitalization and discharge; explanations of medical tests; and information about disease conditions and diverse effects of medications. Such documents differ for each

patient. In addition, future medical care will incorporate personalized medicine, so handing papers containing medical details such as information on diseases or test results to individual patients will become important. For sighted people, documents are actually handed to patients themselves. Blind and partially sighted patients are occasionally accompanied by family members or attendant volunteers who may be present in consultation rooms and also listen to explanations by the medical staff. In Japan the Personal Information Protection Law became fully effective in 2005. This law regards family members as third parties [1] and requires that only patients themselves generally receive medical explanations or documents, except if patients request others to be present with them. This situation suggests the need for medical documents in Braille for blind and partially sighted people. To increase the accessibility of blind and partially sighted patients to medical information, Japan’s Ministry of Welfare issued instructions to Regional Medical Affairs Office on the patient compliance instructions for the disabled including the blind (No. 289, Ministry of Health and Welfare, Healthcare Service Bureau, 19th of August, 1998) and encouraged medical institutions to provide appropriate guidance for drug administration in accordance with the type of disability, for example, using Braille on medical envelopes. This instruction was not compulsory, however. In addition, since fiscal year 2000, Central Social Insurance Medical Council has decided to cover providing medical information in Braille: 10 points (100 yen) for documents and 15 points (150 yen) for booklets, one time per month for each, in the revision of Social Insurance Medical Fee Payment (Ministry of Health and Welfare, Medical Economics Division, Health Insurance Bureau, 29th of February, 2000). At the costs of 100 and 150 yen, however, hospitals have difficulty providing Braille documents. Moreover, medical staff cannot learn Braille in addition to performing their routine duties. We believe that such problems may be solved by having a free automatic Braille translation program available for Braille printing and for users who are unfamiliar with Braille translation. We thus investigated and developed a Web-based Braille translation program that we named eBraille. Braille is a reading and writing system that uses six tactile raised dots. It was invented by a blind Frenchman, Louis Braille, in 1825 [2]. In Japan, Kuraji Ishikawa, a teacher at a school for the blind and speech-impaired in Tokyo, adapted the braille alphabet to Japanese, and in 1890 the school adopted

A. Sugano is with Division of Medical Informatics and Bioinformatics, Kobe University Hospital, Kobe, Japan (phone: +81-78-382-5111; fax: +81-78-382-5839; e-mail: sugano@ med.kobe-u.ac.jp).

M. Ohta is with Division of Medical Informatics and Bioinformatics, Kobe University Hospital, Kobe, Japan (e-mail: ohtam@ med.kobe-u.ac.jp)

M. Ikegami is with Nursing Department, Kobe University Hospital, Kobe, Japan (e-mail: ikegami@ med.kobe-u.ac.jp)

K. Miura is with Division of Medical Informatics and Bioinformatics, Kobe University Graduate School of Medicine, Kobe, Japan (e-mail: 080m812m@stu.kobe-u.ac.jp)

S. Tsukamoto is with Nursing Department, Kobe University Hospital, Kobe, Japan (e-mail: tsukamot@med.kobe-u.ac.jp)

A. Ichinose is with Department of Plastic Surgery, Kobe University Graduate School of Medicine, Kobe, Japan (e-mail: ichise@ med.kobe-u.ac.jp)

T. Ohshima is with Kobe University Graduate School of Medicine, Kobe, Japan (e-mail: ohshima@ med.kobe-u.ac.jp)

E. Maeda is with Division of Medical Informatics and Bioinformatics, Kobe University Hospital, Kobe, Japan (e-mail: emaeda@ med.kobe-u.ac.jp)

M. Matsuura is with Nursing Department, Kobe University Hospital, Kobe, Japan (e-mail: mako@ med.kobe-u.ac.jp)

Y. Takaoka is with Division of Medical Informatics and Bioinformatics, Kobe University Hospital, Kobe, Japan (e-mail: ytakaoka@ med.kobe-u.ac.jp)

his system [3]. The main differences between English braille and Japanese braille are (1) use of indicators which is unique to Japanese language: to express voiced sounds, long vowels or special sounds, and to designate and initiate character systems for the alphabet or signs used in Japanese to differentiate them from *Kana* characters, and (2) text segmentation named *Wakachigaki* [3]. It is a means of writing Japanese in *Kana* characters (phonograms) with spaces between words or phrases (like spacing between words in English and European languages) to make Braille texts easier to understand because such spacing is not used in ordinary Japanese texts [4,5]. The Japanese Braille Committee standardizes such rules for Japanese Braille, which are called Japanese Braille transcription rules [3]. We implemented rules that follows the ones in Japanese Braille transcription rules to develop our translation program named “eBraille” and analyzed its translation accuracy or errors to achieve the higher accuracy. We are also making efforts to provide various medical documents in Japanese braille in our hospital by using eBraille. This paper reports our research development of eBraille and our efforts to provide medical information in Braille in our hospital.

II. MATERIALS AND METHODS

A. Methodologies for the Higher Accuracy

eBraille is a Web-based translation program. Our research developments was as follows: (1) We investigated and analyzed the Japanese Braille transcription rules by The Japanese Braille Committee in order to clarify the linguistic elements which are used for the rules to be implemented on eBraille and how we should use the elements. We then implemented the braille transcription rules on eBraille. (2) For the higher accuracy of eBraille for medical texts, we added medical words to the dictionary for the program. We analyzed word composition in the dictionary and the braille translation accuracy. (3) For the higher accuracy in word segmentation of eBraille, we made statistical learning models (SLMs) by using a generic text chunker called YamCha (Yet Another Multipurpose CHunk Annotator) [6] and analyzed their segmentation accuracy. We used our corpus for evaluation of our program. One is named ordinary text corpus which is based on newspaper articles and the other one is named medical text corpus which is based on nursing records of all treatment departments, medical documents for patients, the Disease List for Specified Rare and Intractable Diseases in the Specified Disease Treatment Research Program of the Japan Ministry of Health, Labour and Welfare [7], and so forth. Those two corpora consist of pairs of ordinary text written in *Kanji* and *Kana* characters and braille text. For the method (1), we used our ordinary text corpus of 5,191 sentences from 233 data set and analyzed the translation accuracy of other braille translation programs for comparison. For the method (2), we used our medical text corpus of 3,180 sentences from 77 data set for the comparison between before and after the dictionary expansion. For the method (3), we made the training and test corpus for the SLMs, from our medical text corpus of total 648 sentences (4233 chunks), because eBraille

showed a relatively lower BTA with them than with other sentences in our medical text corpus. We set the features for the training and test corpus: the resultant morphological analysis which were surface forms of morphemes, *Kana* transcriptions, basic forms, stratified part-of-speech (POS) tags, conjugational forms, and number of linguistic units called morae [8]. The braille translation accuracy (BTA) is calculated as a product of the text segmentation accuracy and the percent correct translation of *Kanji*, correct indicators, and so on. The text segmentation accuracy is the *F*-measure: the harmonic mean of precision (*P*) and recall (*R*) which was calculated as follows:

$$F_{\beta} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

$$\beta = 1 \quad (1)$$

B. Braille brochures for outpatients and inpatients

We made braille brochures for outpatients and inpatients at our university hospital by utilizing our eBraille program. Before translation in braille, we modified the contents so that partially sighted patients could easily find the information which they need first. The examples of such information are necessary procedures for examination or hospital admission. We then itemized or simplified the sentences. Translating the ordinary sentences in *Kanji* and *Kana* characters into braille normally would result in a two- or three-fold increase in the amount of text, similar to what would happen with text after replacement of *Kanji* (ideograms) with *Kana* (phonograms) characters. Therefore, we modified the original sentences by removing repetitive content, omitted the list of Specified Disease Treatment Research Program of the Japan Ministry of Health, Labour and Welfare or patient's bedrooms because we included the contact information for them in addition simplifying the sentences and reduced the text size. Finally we used eBraille to translate the modified original sentences, manually corrected some translation errors, printed with a braille printer (DOG-Multi, Nippon telesoft, Co., Ltd., Tokyo, Japan), and bound brochures with a binder (Docu Bind U-ONE, GBC JAPAN Co., Ltd., Tokyo, Japan).

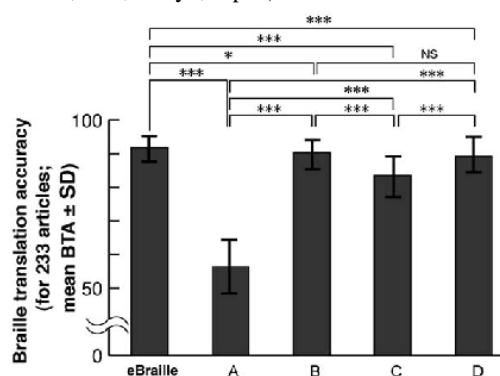


Fig. 1 Braille translation accuracy of translation programs evaluated
p* < 0.05; **p* < 0.001; NS, not significant

III. RESULTS

A. Braille Translation Engine and the BTA

As a result of our analysis of Japanese Braille transcription rules by Japanese Braille Committee, the rules used the following linguistic elements as the basis:

(1) Basic elements for the rules for *Kana* translation

part of speech (POS), the kind of character systems, words in surface form

(2) Basic elements for the rules for text segmentation

POS, the kind of character systems, words in surface form, conjugational form, phonological changes, number of morae [8], the degree of independence for the words, boundary for meaning or pronunciation

In addition, some test segmentation rules required that the elements in (2) were applied to the two consecutive morphemes.

We determined the rules for implementation in accordance with the basic elements and the range of application described above. We excluded the rules with the degree of independence for the words, boundary for meaning or pronunciation because these elements were not clearly defined in the Japanese Braille transcription rules. We then made 111 rules to apply basically to POS and implemented them on our braille translation engine Kobe Univ. Intelligent eBraille Engine for ChaSen

(KUIC) 1.1. We also embedded KUIC to our eBraille program. KUIC is a group of CGI programs which correct *Kana* translation of *Kanji*, segment the text into *Wakachigaki* units, with some correction for braille translation from the output of Japanese morphological analyzer ChaSen [9], and output the segmented *Kana* sentences. The algorithm for text segmentation is as follows:

1. Refer to the POS code of ChaSen for the n^{th} morpheme (initial value for n is 1).

2. Refer to the segmentation rules which is assigned to the POS code.

If a rule(s) is assigned, put two one-byte spaces or no spaces before and/or after the morpheme according to the rule. If no rule is assigned, put one-byte space before and/or after the morpheme in the current position.

3. The current morpheme to which the rule is assigned, are stored in an output buffer.

4. $n \leftarrow n+1$

5. Iterate 1 to 4 until the end of the sentence.

After this process, one-byte spaces between morphemes are converted to two-byte spaces so as to fix the *Wakachigaki* boundaries. The conversion rules are as follows:

(1) If only one one-byte space is put between the morphemes, then delete them.

(2) If there are two to four one-byte spaces, then convert the two to one two-byte space

(3) After the full stop (“.”), exclamation and question marks, put two two-byte spaces regardless of the number of the inserted one-byte spaces. Finally, *Kana* and symbols of indicators and connectors, etc. are converted into braille. Each character is converted into

TABLE I
COMPARISON OF ACCURACIES BEFORE AND AFTER DICTIONARY
EXPANSION

Score	Before expansion	After expansion
Text Segmentation (F_1)	92.73 ± 5.64	96.06 ± 2.74
Kana or Symbols translation (%)	94.77 ± 4.07	96.32 ± 2.62
Braille Translation Accuracy (BTA)	88.06 ± 8.29	92.57 ± 4.60

TABLE II
CHARACTERISTICS OF THE EBRILLE PROGRAMS

	eBraille	eBraille-M	eBraille-TM	eBraille for English
Japanese braille translation	○	○	○	—
English braille translation	—	—	—	○
original dictionary (ipadic 2.7.0, 239,631 words)	○	○	○	—
Medical words dictionary (8,170 words)	—	○	—	—
Oriental Traditional Medicine dictionary (17,290 words)	—	—	○	—

TABLE III
SEGMENTATION ACCURACY FOR MEDICAL TEXT

	Precision (%)	Recall (%)	F-measure (F_1)
eBraille-M	93.43	92.84	93.13
SLM	93.49	93.23	93.34
Combined Model	94.57	94.49	94.53

TABLE IV
SEGMENTATION ACCURACY FOR COMPOUND NOUNS

	Precision (%)	Recall (%)	F-measure (F_1)
eBraille-M	89.13	88.57	88.83
SLM	92.51	93.67	93.08
Combined Model	91.34	92.54	91.92

six digit numbers, which consist of 0 or 1. The six digits designate the six dot pattern of the braille that follows North America Braille Computer Code (NABCC). After obtained the dot patterns, the numbers are converted to pbm format to produce GIF images for tactile graphics embossers and braille styluses and slates, or converted to ASCII codes to create a BASE format file for braille printers. We then evaluated eBraille with KUIC 1.1 and other braille translation programs by using our ordinary text corpus for comparison. The BTA scores are statistically analyzed by using repeated measures ANOVA with post-test Tukey. As a result, eBraille showed the significantly highest BTA (Fig 1).

B. Expanded Dictionary and SLM

To increase the BTA of eBraille for medical text, we added 8,170 medical words to the dictionary and evaluated eBraille with the expanded dictionary. Our result showed the higher BTA than that of eBraille with the original dictionary (Table I). Our analysis of relationship between the dictionary composition and BTA suggested that optimal dictionary composition for the texts for translation would contribute the higher BTA. Thus we developed eBraille-M for medical text translation, eBraille-TM for the literature of Japanese Traditional Medicine with an additional 17,290 technical terms in the field (Table II). In addition, we also developed eBraille-E for texts in English into English braille. All these braille translation programs are open on the Internet (<http://ebraille.med.kobe-u.ac.jp/>). For the purpose to achieve higher accuracy of the text segmentation for braille, we made SLMs and analyzed their segmentation accuracy F_1 to compare with that of eBraille-M. As a result, the SLM showed a higher segmentation accuracy than did eBraille-M: eBraille-M showed 93.13 and SLM showed 93.34. Comparison of segmentation errors showed that the SLM output the correct segmentation of the compound nouns in Chinese characters such as 視野■欠損 (“visual field loss”) and 歯科■口腔 (“dental and oral”), where ■ designates a space. However, it failed to divide compound words and phrases such as ビタミン■D (“vitamin D”), CT■にて (“by CT”), and also failed to output the phrase into one segment such as 吸引にて (“by suction”), which the eBraille-M translation engine KUIC output correctly. We then compared the two programs for their text

segmentation accuracy of compound nouns because SLM output compound nouns correctly. The F_1 value of eBraille-M was 88.83 and that of SLM was 93.08; this difference was larger than that for our medical text corpus. These results indicated that the SLM effectively segmented compound nouns but failed segmentation of words and phrases described above, which KUIC never failed in accordance with the implemented Braille transcription rules.

Next we combined the SLM with the rule-based eBraille-M and compared its accuracy with the two programs, that are SLM and eBraille-M. The combined model was a new statistical model with the additional new feature, eBraille-M output. As a result of the evaluation of the three programs, the combined model showed the highest F_1 value (Table III). However, the SLM had the highest F_1 value for compound nouns (Table IV). We then compared the segmentation errors of SLM and the combined model. The comparison showed that the error patterns of the two models differed. The combined model showed the same output as did eBraille-M, namely according to the rules implemented with KUIC, while SLM failed such outputs. The combined model failed the segmentation at the second mora of compound nouns, whereas the SLM output was correct. In addition, the errors of the combined model were the same as that of eBraille-M. Thus our result indicated that the combined model could learn the correct outputs of eBraille-M and also could learn the errors of eBraille-M.

C. Providing Braille Brochures in our Hospital

The hospital brochures for outpatients and inpatients were printed both in braille and ordinary text in *Kanji* and *Kana* characters so that hospital staff could understand the content (Fig 2). The braille brochures were placed at the general information counter in our hospital, general reception window, and reception window in the Division of Ophthalmology to respond to requests from patients. They were also placed in the information center for patients and on the bulletin board on the aisle for patients to touch and read freely. According to the partially sighted patients, these braille brochures were welcomed and evaluated favorably as part of the hospital's efforts to improve accessibility.

IV. DISCUSSION AND CONCLUSION

With the development of the braille translation engine KUIC, our Web-based Braille program, achieved a high BTA score, which can be for the practical use, for the ordinary text such as newspaper articles. In addition, eBraille-M with the expanded dictionary with additional medical words showed the higher BTA than that of eBraille, which use the original dictionary. Our experiments showed that the combined model, which was constructed from eBraille-M and SLM showed the highest text segmentation accuracy (F_1 value) for medical text. However, SLM had the higher segmentation accuracy for compound nouns than that of the combined model. These results suggested that use of both combined model and SLM would contribute the higher accuracy. Meanwhile, eBraille-M could have high F_1 value more than 95.0, depending the kind of medical texts (data

not shown). So we will analyze the segmentation accuracy by the experiments with the larger size of training and test corpus.

With our research development, eBraille program has been improved its BTA score for medical translation and we started providing braille documents for the blind and partially sighted patients. We will continue to research and develop the system to improve the accessibility for the blind and partially sighted patients both in the information technology and organizational aspects. We announced that our hospital could provide medical information in braille and gave the correspondence to response the requests from the medical departments in the meeting for the directors or the nurse heads. In the future, we hope that our eBraille program is cooperated with our hospital information system in order to translate the medical documents such as the treatment schedule or informed consent in braille as needed. Moreover, we will work on the supportive organization such as Braille libraries and Nippon Lighthouse Welfare Center for the Blind, to keep them informed about our program.

APPENDIX

We would like to express our sincere condolences for the sufferers and their families of Tōhoku earthquake and tsunami on Friday, 11 March 2011. This earthquake disaster made us realize how important it is to provide necessary information. Thus we put up the portal site (<http://suzume.med.kobe-u.ac.jp/eq/>) which provides the information for the sufferers on the same day (Fig 2): the status of medical institution, hospitals which provide dialysis or insulin, announcement from banks, nuclear power plant disaster in Fukushima pref., etc. Such information are available in BASE format for braille printers or PDF for tactile graphics embossers. We will continue to provide information.



Fig. 2 Hospital brochures in braille and the example of printed page

ACKNOWLEDGMENT

This work was supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) (No.

101707012) of the Ministry of Internal Affairs and Communications, Japan; a Grant-in-Aid for Exploratory Research (grant no. 19659563), Grant-in-Aid for Scientific Research (C) (grant no. 21592692 and 23593135) and a Grant-in-Aid for Young Scientists (B) (Grant no. 21790497) from the Japan Society for the Promotion of Science.

REFERENCES

- [1] Act on the Protection of Personal Information (Japan Law No.57, 2003). (in Japanese) URL: <http://law.e-gov.go.jp/htmldata/H15/H15HO057.html> [accessed 2011 May 9]
- [2] J. Jiménez, J. Olea, Torres J, I. Alonso, D. Harder, K. Fischer. Biography of louis braille and invention of the braille alphabet. *Surv Ophthalmol* 2009; 54(1): 142-9
- [3] Japanese Braille Committee, Japanese Braille Description Rules, 2001. Tokyo: Daikatsuji Co. Ltd.; 2001 (in Japanese). ISBN:978-4860550004
- [4] Kadota M. Japanese Braille Tutorial. 1997; URL: http://www.duxbury.com/Japanese_Braille_Tutorial.pdf [accessed 2011 March 18]
- [5] Unger JM. Japanese Braille. *Visible Language* 1984; 18(3):254-266.
- [6] Kudo T. YamCha: Yet Another Multipurpose CHunk Annotator. URL: <http://chasen.org/~taku/software/YamCha/> [accessed 2011 May 9]
- [7] Japan Intractable Diseases Information Center. URL: <http://www.nanbyou.or.jp/english/> [accessed 2011 June 9]
- [8] Trubetzkoy NS. *Grundzüge der Phonologie* (Principles of Phonology). Berkeley, CA: University California Press; 1958/1969.
- [9] Matsumoto Y, Kitauchi A, Yamashita T, Hirano Y. Japanese Morphological Analysis System ChaSen Version 2.0 Manual. Information Science Technical Report TR99009, Nara Institute of Science and Technology. URL: <http://library.naist.jp/mylmedial/dllmedial/show.cgi?bookid=23003> [accessed 2011 May 25]



Fig. 3 Our portal site for the sufferers of Tōhoku earthquake and tsunami