JAPANESE-ENGLISH CODE-SWITCHING SPEECH DATA CONSTRUCTION

Sahoko Nakayama, Takatomo Kano, Quoc Truong Do, Sakriani Sakti, and Satoshi Nakamura

Augmented Human Communication Laboratory
Graduate School of Information Science
Nara Institute of Science and Technology, Japan

ABSTRACT

As the number of Japanese-English bilingual speakers continues to increase, code-switching phenomena also happen more frequently. The units and locations of switches may vary widely from single word switches to whole phrases (beyond the length of the loanword units). Therefore, speech recognition systems must be developed that can handle not only Japanese or English but also Japanese-English code-switching. Consequently, a large-scale code-switching speech database is required for model training. But collecting natural conversation dialogues of Japanese-English data is both time-consuming and expensive. This paper presents the construction of Japanese-English code-switching speech data by utilizing a Japanese and English text-to-speech system from a bilingual speaker. Various switching units are also investigated including units of words and phrases. As a result, we successfully constructed over 280-k speech utterances of Japanese-English code-switching.

Index Terms— Data construction, code-switching, Japanese and English languages

1. INTRODUCTION

Code-switching, which is defined as an alternative linguistic variety between two languages in the same conversation, usually within the same conversational turn or even in the same sentence of that turn [1], has been studied for several decades. Most researchers agree that it plays a vital role in bilingualism and is more than a random phenomenon [2]. Several studies on code-switching speech recognition also exist. Since the units and locations of switches may vary widely from single word switches to whole phrases (beyond the length of standard loanword units), simply utilizing multiple monolingual ASR systems is difficult. Furthermore, unexpected switching between languages greatly complicates the recognition of such speech.

White et al.[3] investigated alternatives to model the acoustics for multilingual code-switching, and Imseng et al. [4] proposed an approach to estimate the universal phoneme posterior probabilities for mixed language speech recognition. Vu et al. focused on the speech recognition of Chinese and English code-switching [5] and proposed approaches for phone merging in combination with discriminative training as well as the integration of language identification systems in decoding processes. Recently, Yilmaz et al. investigated the impact of bilingual deep neural networks in the contexts of Frisian and Dutch code-switching [6].

Despite extensive studies on code-switching speech recognition in bilingual communities, the Japanese-English case has received scant research up to now because the Japanese language shows almost complete dominance over the country's entire population. Furthermore, a large number of its loanwords, most of which have been borrowed from English, are already commonly used [7, 8].

In recent years, the number of bilingual speakers has increased in Japan. According to a Ministry of Health, Labour and Welfare (MHLW) survey, the number of children who have a foreigner parent in Japan has increased from 13,686 in 1990 to 21,180 in 2016 [9]. The Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) also reported that the number of school-age children (from elementary to high school) who have returned from living abroad increased from 5,900 in 1977 to 12,527 in 2015 [10]. Moreover, the number of Japanese students who study abroad has increased from 23,633 to 60,643 in the past ten years, as reported by the Japan Student Services Organization (JASSO) [11]. Under this background, the phenomenon of Japanese-English code-switching is also more frequent. Code-switching often happens unconsciously; speakers may be unaware of it.

Nakamura [12] surveyed the code-switching of a Japanese child who came to the United States by recording 30-minute segments of conversation between the subject and his mother and found that mixed turns (both English and Japanese in a single turn) occurred 59 times. In other words, he started his utterances in Japanese and changed to English in the middle of a single turn. Fotos also analyzed four hours of conversation of four bilingual children living in Japan with American parents or one American parent and observed 153 instances of code-switching [13].

The above reports reveal that some people actually use code-switching in everyday life. Therefore, speech recognition systems must be developed that can handle not only

Japanese or English language but also Japanese-English code-switching. Consequently, models must be trained with code-switching speech. In addition, a significant amount of data is necessary for training a deep neural network model. Unfortunately, no such large-scale Japanese-English code-switching speech database exists for model training. Unfortunately, collecting natural conversation dialogues of Japanese-English data is time-consuming and expensive. This paper describes our Japanese-English code-switching speech database that utilizes a Japanese and English text-to-speech (TTS) system from a bilingual speaker. The target of the code-switching speaker is a native Japanese speaker who can also speak English.

2. CODE-SWITCHING IN JAPANESE

Code-switching is classified into two primary categories: inter-sentential and intra-sentential. Inter-sentential code-switching is when the language switch is done at the sentence boundaries. In intra-sentential code-switching, the shift is done in the middle of a sentence. Intra-sentential code-switching has different types of code-switching based on locations and lengths. Intra-sentential code-switching can occur from the length of just a single word unit to an entire phrase.

We differentiate between code-switching words and loanwords (borrowings), which are generally used by monolingual speakers in monolingual societies. For example, meritto is borrowed from the English word merit, but Japanese monolinguals use it in their monolingual society. Thus, it is not counted as a switch. Below are examples of code-switching and loanword insertion that were reported in existing studies. Regarding loanword and intra-sentential phrase insertion code-switching, here is a quotation from content that was actually spoken in the lecture transcription at our institute:

- [Loanword insertion]:
  `tyukangengo wo tsukatta toki no meritto ni naniga aruka?`
  (*What is the merit of using an interlingua?*)

- [Intra-sentential word-level code-switching]:
  Trust-`shiteru hito ni dake kashiteageru no.`
  (*I lend (it) only to a person I trust.*) [12]

- [Intra-sentential phrase insertion code-switching]:
  "He reckons the current account deficit" `to yuu bunga aruto suruto` "He" `to yuuno ga ninsyoudaimeishi dearu.`
  (*If there is a sentence "He determines the current account deficit," "he" is a personal pronoun.*)

- [Inter-sentential code-switching]:
  `Aa, soo datte nee.` On the honeymoon, they bought this.
  (*Oh, year, you're right. On their honeymoon, they bought this.*) [12]

3. AVAILABLE DATA RESOURCES

3.1. Monolingual BTEC Text Data

The ATR Basic Travel Expression Corpus (BTEC) [14, 15] covers basic conversations in travel domains, such as sightseeing, restaurants, hotels, etc. Its sentences were collected by bilingual travel experts from Japanese/English sentence pairs in travel domain phrasebooks. The BTEC has been translated into French, German, Italian, Chinese, and Korean. In this study, we used Japanese-English BTEC text data called BTEC 1, 2, and 3. Table1 lists their basic statistics.

Table 1. Basic statistics of BTEC text data

	BTEC 1	**BTEC 2**	**BTEC 3**	**BTEC 4**
# Sentences	172k	46k	198k	74k
# Word tokens	1,174k	341k	1,434k	548k
# Word types	28k	20k	43k	22k

3.2. Bilingual BTEC Speech Data

We also utilized the Bilingual BTEC speech data that were originally constructed to emphasize a speech translation study [16]. 1015 Japanese-English sentence pairs were selected from 16,000 BTEC sentences. The recording was done with three bilingual speakers in a quiet environment, and the WAV audio was recorded with a frequency of 16 KHz, 16 bits, and a single channel. After the recording, all the audio files were verified to ensure that no clipping was caused by a speaker who was talking too loudly. In this study, we only use the data from one bilingual speaker. A transcription sample is shown in Table 2.

Table 2. Sample of bilingual BTEC speech data

Language	**Transcription**

Japanese	`Kamera desu.`
	`Suteki na hi ne?`

English	That's a camera. Beautiful day, isn't it?

4. ENGLISH AND JAPANESE HMM-BASED SPEECH SYNTHESIS

Since large-size BTEC speech utterances for the corresponding BTEC 1-4 text corpora are unavailable, we developed monolingual Japanese and English TTS systems. To enable code-switching speech from the same speaker, both TTS systems were generated based on a small set of bilingual BTEC speech data described above and conducted using an open-source speech synthesis engine called a HMM-based Speech Synthesis System (HTS) [17].

A total of 960 training utterances were used for both the Japanese and English TTS systems. The only difference between them was the data; the model architecture was identical. The speech signals were sampled at 16 kHz and windowed using a 25-ms Hamming window and a 5-ms frame shift. The feature vector consisted of 40-dimensional Mel-Generalized cepstral coefficients mgc, a log f0 (lf0), and band aperiodicities (bap). We used the WORLD vocoder [18] to generate the speech.

5. CODE-SWITCHING DATA CONSTRUCTION

An overview of the data construction is illustrated in Fig. 1 and consists of two main processes: text and speech data construction. Both processes are described in more detail below.

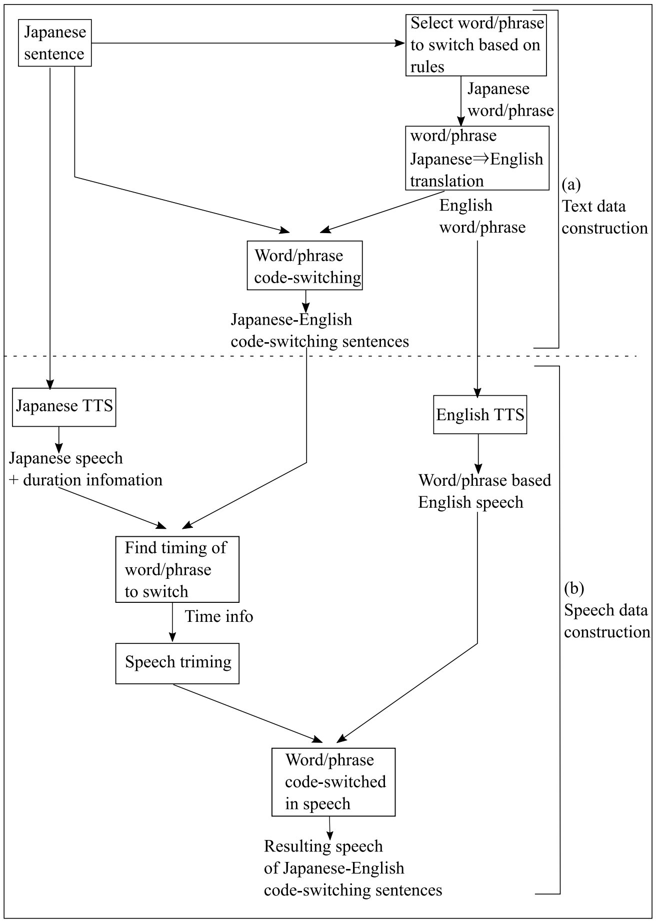
Japanese sentence
Select word/phrase to switch based on rules
Japanese word/phrase
word/phrase Japanese⇒English translation
English word/phrase
(a) Text data construction
Word/phrase code-switching
Japanese-English code-switching sentences
Japanese TTS
English TTS
Japanese speech + duration infomation
Word/phrase based English speech
Find timing of word/phrase to switch
Time info
Speech triming
(b) Speech data construction
Word/phrase code-switched in speech
Resulting speech of Japanese-English code-switching sentences

Fig. 1. Overview of Japanese-English code-switching data construction

5.1. Text Data Construction

5.1.1. Intra-sentential word-level code-switching

We first selected words from the BTEC text data that are commonly written with Japanese katakana characters and translated them into English using Google translation API. Intra-sentential word-level code-switching sentences were created by inserting the translated English words into the original Japanese sentences.

Note that although Japanese katakana is often used to describe loanwords, since we later used the speech pronounced by a native English speaker in an English TTS system to generate utterances, they can be distinguished from loanwords.

Here, code-switching can be done by inserting a single word or a chunk of two-to-three words. Two or more words may also be inserted in different places within one sentence. Below is an example of the resulting intra-sentential word-level code-switching sentences:

- Kanko basu no pamphlet wa ari masu ka?
(*Do you have any brochures for the sightseeing bus?*)

5.1.2. Intra-sentential phrase-level code-switching

In contrast with the word-level code-switching case, we selected phrases from the BTEC text data beyond the length of the loanword units. Here, to produce natural conversations, we referred to the actual examples that were reported in existing studies [12], especially a switching pattern from Japanese-to-English phrases after Japanese particles appear.

We translated the Japanese phrases after Japanese particles into English using the Google Translate API and re-inserted them into the original sentences. We used the following Japanese particles:

ga, wo, ni, he(e), to, kara, yori, ba, temo, keredo, noni, node, kara, nari, nagara, tari, and tsutsu.

Below is an example of the resulting intra-sentential phrase-level code-switching sentences:

- gasu iri no tansan sui wo bring me with two ice.
(*Bring me two bottles of carbonic minerals and some ice, please.*)

5.2. Speech Data Construction

5.2.1. Synthesizing Speech

Given intra-sentential code-switching sentences, we generate Japanese and English speech using our monolingual TTS systems. Japanese speech utterances were synthesized from Japanese sentences, and files with the duration information of Japanese speech were simultaneously created. English speech utterances were also synthesized from translated English words or phrases.

5.2.2. *Trimming and Concatenation*

Based on the obtained time information, we trimmed the Japanese speech utterances at the switching position using the sox[1] command. Next we concatenated them with the English speech part also using a sox command (Fig. 2). As a result, we created speech of Japanese-English code-switching sentences.

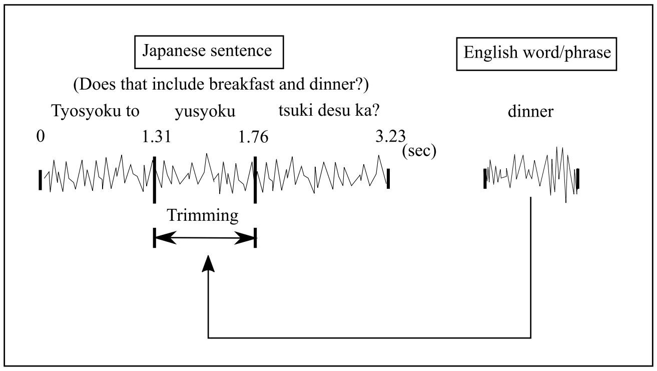

Fig. 2. Speech trimming and concatenation process

6. DATA ANALYSIS

After constructing the data, we investigated what percentage of English and Japanese are included in the code-switching by counting the number of words. A morpheme analyzed by MeCab[2] is counted as one unit. We counted the number of morphemes in each language and the number of whole morphemes and calculated the ratio. Morphemes that are neither English nor Japanese words (such as numbers and symbols) were excluded from the results because they were less than 0.15%. The investigation results are shown in Table 3. In the monolingual Japanese corpus, the proportion of Japanese is 100%, and the proportion of English is 100% in the monolingual English corpus. But in the word insertion, the proportion of Japanese words is 88% and only 12% English words; in the phrase insertion, the proportion of Japanese words is 46% and 56% are English words.

	Utterances	Japanese words [%]
Monolingual Japanese	273k	100%
Japanese sentences with English words	146k	88%
Japanese sentences with English phrases	146k	46%
Monolingual English	273k	0%

Table 3. Statistics of Japanese-English code-switching speech utterances

7. CONCLUSION

In this paper, we described the construction of a Japanese-English code-switching corpus. Japanese-English code-switching speech data were generated by utilizing Japanese and English TTS systems from a bilingual speaker. Various units of switches were also investigated including units of words and phrases. The resulting corpus includes intra-sentential word insertion code-switching with 146-k speech utterances (12% English words) and intra-sentential phrase insertion code-switching with 146-k speech utterances (54% English words).

In the future, we will further investigate the quality of code-switched synthesized speech in comparison to natural speech and utilize it to enhance our speech recognition system for the bilingual community.

8. ACKNOWLEDGEMENT

Part of this work was supported by JSPS KAKENHI Grant Numbers JP17H06101 and JP17K00237.

9. REFERENCES

[1] C. Myers-Scotton, *Social motivations for codeswitching. Evidence from Africa.*, Oxford: Clarendon Press, 1993.

[2] Jeff McSwan, “The architecture of the bilingual language faculty: Evidence from intrasentential code switching,” *Bilingualism: Language and Cognition*, vol. 3, no. 1, pp. 37–54, 2000.

[3] C. White, S. Khudanpur, and J. Baker, “An investigation of acoustic models for multilingual code switching,” in *INTERSPEECH*, 2008, pp. 2691–2694.

[4] D. Imseng, H. Bourlard, M. Magimai-Doss, and J. Dines, “Language dependent universal phoneme posterior estimation for mixed language speech recognition,” in *ICASSP*, 2011, pp. 5012–5015.

[5] N. T. Vu, D. C. Lyu, J. Weiner, D. Telaar, T. Schlippe, F. Blaicher, E. S. Chng, T. Schultz, and H. Li, “A first speech recognition system for Mandarin-English codeswitch conversational speech,” in *ICASSP*, 2012, pp. 4889–4892.

[1] Sox – http://sox.sourceforge.net/

[2] MeCab is a morphological analyzer developed by Kyoto University.

[6] E. Yilmaz, H. den Heuvel, and D. van Leeuwen, "Investigating bilingual deep neural networks for automatic recognition of code-switching Frisian speech," *Procedia Computer Science*, vol. 81, pp. 159 – 166, 2016, SLTU-2016 5th Workshop on Spoken Language Technologies for Under-resourced languages.

[7] J.C. Maher and K. Yashiro, *Multilingual Japan*, chapter Multilingual Japan: An introduction, pp. 1–17, 1995.

[8] Y. Morishima, "Conversational code-switching among japanese-english bilinguals who have japanese background," M.S. thesis, Edith Cowan University, 1999.

[9] Japanese Ministry of Health, Labour and Welfare, "Overview of the population statistics in 2016 [in Japanese]," *http://www.mhlw.go.jp/*, 2016.

[10] Japanese Ministry of Education, Culture, Sports, Science, and Technology, "School basic survey in 2015 [in Japanese]," *http://www.mext.go.jp/*, 2015.

[11] Japan Student Service Organization, "Survey on japanese student abroad situation in 2016 [in Japanese]," *http://www.jasso.go.jp/*, 2016.

[12] M. Nakamura, "Developing codeswitching patterns of a Japanese/English bilingual child," in *Proceedings of the 4th International Symposium on Bilingualism*, 2005, pp. 1679–1689.

[13] S.F. Fotos, "Japanese-English code switching in bilingual children," *JALT Journal*, vol. 12, no. 1, pp. 75–98, 1990.

[14] T. Takazawa, G. Kikui, M. Mizushima, and E. Sumita, "Multilingual spoken language corpus development for communication research," *The Association for Computational Linguistics and Chinese Language Processing*, vol. 12, no. 3, pp. 303–324, 2007.

[15] G. Kikui, E. Sumita, T. Takezawa, and S. Yamamoto, "Creating corpora for speech-to-speech translation," in *EUROSPEECH*, 2003, pp. 381–384.

[16] Q.T. Do, S. Sakti, G. Neubig, T. Toda, and S. Nakamura, "Collection and analysis of a japanese-english emphasized speech corpus," in *Proceedings of Oriental COCOSDA*, 2014.

[17] H. Zen, T. Nose, J. Yamagishi, S. Sako, T. Masuko, A. Black, and K. Tokuda, "The HMM-based speech synthesissystem (HTS) version 2.0," in *ISCA Workshop on Speech Synthesis*, 2007, pp. 294–299.

[18] M. Morise, F. Yokomori, and K. Ozawa, "WORLD: A vocoder-based high-quality speech synthesis system for real-time applications," *IEICE Transaction on Information and Systems*, vol. 99, no. 7, pp. 1877–1884, 2016.