Interactive Lyric Translation System: Implementation and Experiments

Ayano Nishimura
Ochanomizu University
Tokyo, Japan
nishimura.ayano@is.ocha.ac.jp

Takayuki Itoh
Ochanomizu University
Tokyo, Japan
itot@ocha.ac.jp

ABSTRACT

We can listen to music from many different countries due to the evolution of the Internet. However, understanding lyrics written in foreign languages is still difficult, even though many international songs have been translated.

In this paper, we propose an interactive lyric translation system and describe its implementation. Users can modify lyrics by selecting a sample lyric from candidate translations and then freely edit the lyric using the proposed system. The proposed system also allows users to listen to their translation by applying singing voice synthesizer and search for related words. We conducted experiments with 12 participants to compare the lyric translation by the proposed system to manual lyric translation. The translation using the proposed system had better evaluation results comparing to the manual translation.

1 INTRODUCTION

The Internet has made it possible to listen to and enjoy international music and access the lyrics written in foreign languages at any time. However, the lyrics written in a foreign language are often difficult to interpret.

Meanwhile, many international songs have been translated while fitting original melodies. Effective translation is still a professional task because it requires understanding the original

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313046

KEYWORDS

Lyric Translations; Music; User Interface.

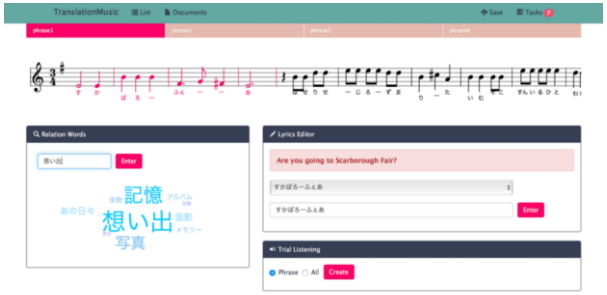

Figure 1. Lyric Translation System.

language, musical phrases, rhythm, and pronunciation. There have been studies that the language of lyrics makes a difference in the song rhythm [6], [7].

To address these issues, we have developed an interactive lyric translation system to translate lyrics written in foreign languages into native languages of users (English to Japanese in our implementation), as shown in Figure 1. The proposed system displays an automatically translated lyric and its musical score for editing according to the preferences of users. This paper describes the proposed system and discusses the experimental results. We evaluated the proposed system experimentally with 12 participants.

2 RELATED WORK

Studies into lyric generation and can be roughly divided into automatic and interactive techniques. “Tra-la-lyrics 2.0” automatically generates song lyrics in Portuguese based on a melody [3]. Our study aims at developing an interactive system, while the system aims automatic translation, featuring an algorithm to modify melody according to lyrics and generate temporary translations.

“Hafez” is an interactive poetry generation system without a melody that incorporate Recurrent Neural Network with a Finite State Acceptor [2]. The “pâtisier” lyric creation support system suggests lyrics that suit the number of morae and vowels [1]. “LyriSys” is a lyric creation support system that considers the comprehensive structure of lyrics, such as verses and choruses [9].

Meanwhile, the task of reducing the number of morae in the translated lyrics is related to sentence compression in natural language processing. There have been short and simple sentence generation methods such as an electrical bulletin board in the Shinkansen trains by focusing on the end of the sentence [10].

3 LYRIC TRANSLATION ALGORITHM

Herein, a mora is defined as a vowel, a consonant followed by a vowel or contracted sound. The Japanese syllabic nasal is counted as 0 mora in this time. Contracted sounds are indicated when certain “i” vowel characters are followed by smaller sized “ya(ゃ),” “yu(ゅ),” and “yo(ょ).” Translation from English into Japanese requires dividing the melodies because the number of morae usually increases. The meaning of the original lyric must be abbreviated [4] to prevent the increase in the number of morae.

This processing procedure is shown in Figure 2. The procedure is roughly divided into the following three parts: (1) translation of source language, (2) lyric processing of translated sentences, and (3) matching generated lyrics to the melody.

3.1 Translation of Source Language

Our implementation applies Microsoft Translator Text API [12] to get translated sentences from English lyrics described in MusicXML. Periods, commas, colons and semicolons delimiters are regarded as one phrase. The number of musical notes is taken as the number of lyric morae. This process acquires translated sentences and the number of musical notes of each phrase.

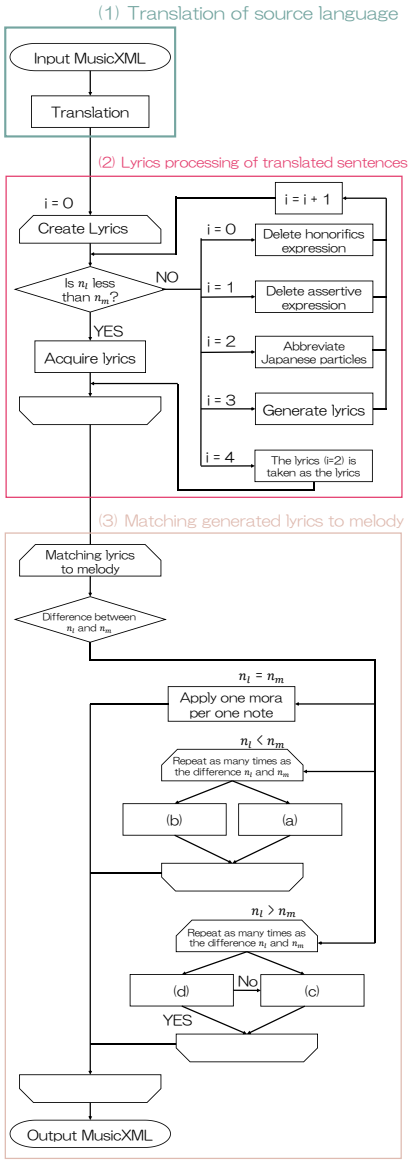

Figure 2. Processing Procedure of the Lyric Translation.

3.2 Lyric Processing of Translated Sentence

This process generates lyrics from the acquired machine translation sentences and the number of musical notes. We define the number of lyric morae n_l and the number of musical notes n_m. The following four operations are executed one by one when the number of morae in a machine translation sentence of one phrase exceeds the number of musical notes: **i=0**: delete honorific expression, **i=1**: delete assertive expression, **i=2**: abbreviate Japanese particles, **i=3**: generate lyrics.

i=3 is executed when the number of morae exceeds the number of musical notes after processing i=0 to i=2. i=3 applies 2-gram to generate the lyric so that the number of morae gets equal to the number of musical notes. Note that a word which has the highest tf-idf value in a single phrase is given as the first word. tf-idf is a statistical value for evaluating the importance of words included in a document. We used the corpus created from Project Gutenberg and Aozora Bunko as a document to calculate tf-idf value [8]. We also used the Japanese Wikipedia corpus to generate lyrics [11].

It is sometimes difficult to generate the lyric in i=3 depending on the number of morae and given words. In this case, the system switch **i=4**, and applies the sentence modified by i=0 to i=2 as the lyrics and allows to exceed the number of musical notes.

3.3 Matching generated Lyrics to Melody

We divide the relationship between n_l and n_m as follows. Case 1: $n_l = n_m$, Case 2: $n_l < n_m$, and Case 3: $n_l > n_m$. The generator applies either of the two algorithms for matching melody and lyric for Cases 2 and 3. We do not need to modify the melody if n_l and n_m satisfies Case 1.

Japanese singing usually allows to stretch a vowel of one character and sing a plurality of notes. In Case 2, the number of lyric morae are increased by inserting stretch bar in the k-th ($k > 0$) musical note in a phrase at Figure 2-(a), when the k-th musical note duration from the beginning the phrase is dur(k) < dur(k + 1). Note that we define dur() is a function of sound duration. If all the musical note durations are equivalent, the algorithm inserts stretch bar at the end of the lyric. The second algorithm merges the shortest note in the phrase with the previous note at Figure 2-(b). If it is impossible, the algorithm merges the shortest note with the next note.

In Case 3, one of the longest musical notes that is the closest to the beginning of the phrase is divided into two notes which have the half duration. Or, the note is divided into two to one if it is a dotted note (Figure 2-(c)). The second algorithm defines the minimum note duration in advance, and notes are divided so as not to be shorter than the minimum duration (Figure 2-(d)). Our implementation defines the eighth note as the minimum note duration. Even if all notes are divided by the minimum duration, if n_l gets larger than n_m, the other algorithm is then applied.

4 IMPLEMENTATION

We implemented the system as a Web application (Figure 3). As shown, the user interface is divided into two parts: (1) the score, and (2) edit components.

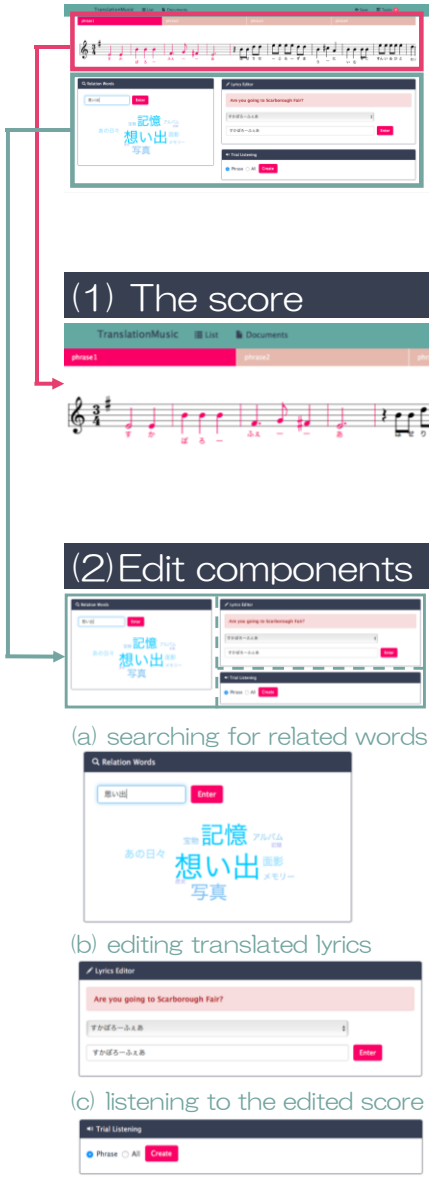

Figure 3. System Interface.

The score component has a phrase bar and a score. This phrase bar that shows the phrase being edited is displayed at the top of the score component, and the target phrase is shown in pink in the score. Users can change phrases by scrolling through a score or clicking on the phrase number.

The edit component facilitates the following actions: (a) searching for related words, (b) editing translated lyrics, and (c) listening to the edited score.

4.1 Searching for Related Words

The proposed system applies a word cloud to display related words, as shown in Figure 3 (2)-(a). Related words determined by the cosine similarity are displayed when a user inputs a word on the text field. Words which have a higher similarity with the input word are displayed by larger letters. Note that display words morae are not equal to inputted word morae.

4.2 Editing Translated Lyrics

This function comprises a panel, a selection widget, and a text field (Figure 3 (2)-(b)). The panel shows the original lyrics so that users can compare edited lyrics to the original lyrics.

The selection widget with four selectable lyric translation candidates. The current system displays automatically-translated lyrics as the initial candidate, whereas translated lyrics as the second to fourth candidates. Since the lyric translation algorithm takes a long time if enter process of lyrics generation, user interface uses lyrics data output in advance. The words in the machine-translated sentences with the highest tf-idf values are shown at the beginning of all the lyrics. The text field shows the lyrics selected from the selection widget, and users can edit this text freely.

Melody notes are automatically divided when the number of morae in the input lyric is beyond the number of musical notes. The system randomly applies one of two algorithms for note division, shown in section 3.3.

4.3 Listening to Edited Scores

This function allows users to listen to their edited songs, is useful for users that cannot read musical scores. Users can select the listening target as either a single phrase or the entire song, as shown in Figure 3 (2)-(c). We apply a singing voice synthesizer “Sinsy” [5]. Sinsy generates singing voices based on music scores written in MusicXML applying Hidden Markov Model and Deep Neural Network.

5 COMPARATIVE EXPERIMENTS

We conducted experiments which compare lyric translations by the presented system and manual lyric translations to clarify which method is the better lyric translation method. In each case, the participants could continue lyric translation tasks until they were satisfied with the translation. The following describes the experimental procedures.

Following is the procedure of the lyrics translation using the proposed system; (1) Listen to the original song and see the original lyrics. (2) Experience with manual lyric translations (five minutes). (3) Explain how to use the proposed system. (4) Practice using the proposed system (five minutes). (5) Translate the lyrics by using the proposed system. (6) Answer a questionnaire.

Participant Comments About the System:

Positive comments:

- It was helpful to listen to the edited song in real time (two participants).
- Lyric editing using translation candidates is useful (four participants).
- Automatic change of the note division is good.

Negative comments:

- I could not change note division by myself (three participants).
- It seems there are no rule for musical note division because one of the algorithms is randomly applied.
- I wanted a function to listen to the original and edited songs for comparison because I had to listen to the original song on YouTube while working.

Participant Comments About Manual Translation:

Positive comments:

- It was easy to rewrite lyrics.
- Lyric translation could be edited according to the English lyrics.
- It was easy to understand where the lyrics were edited.
- I could freely modify note division.
- It is good that Chinese characters can be used.

Negative comments:

- It was troublesome to rewrite lyrics.
- It was difficult to imagine the sound of the edited song (two participants).
- I could not match notes and lyrics.

Following is for manual lyric translations; (1) Listen to the original song and see the original lyrics. (2) Explain how to use the proposed system. (3) Experience to use the proposed system (five minutes). (4) Perform manual lyric translations. (5) Answer a questionnaire.

We used the score of "Scarborough Fair" in the experiment. Also, we used "Lavender's Blue" as a score while demonstrating manual lyric translations and experiencing with the proposed system. These songs are English nursery rhyme. The participants provided a relative evaluation of lyric translations using the proposed system and manual translations in both experiments.

6 RESULTS

We invited only participants who could read musical score for these comparative experiments. We asked six participants to translate using the proposed system, and the other six participants performed manual lyric translations. Nine participants were female and other three were male (21 to 54 years; average: 29 years). We asked participants to answer as five-point Likert scale on the questionnaire. Participants gave the comments about the system and the manual translation process, on a sidebar.

6.1 Result of Comparison

We discuss the results of the comparative questionnaire, in which similar questions were asked for both the proposed system and the manual translations task. Figure 4 shows the results, including t-test, the average and the standard error for the questionnaire. The proposed system received higher evaluations for all the evaluation items. In addition, there were significant differences for three of the four items.

The degree of satisfaction with the completed songs showed an obvious difference relative to the standard error. The participants could edit steadily and were satisfied with the song generated using the proposed system. On the other hand, several participants gave up translating lyrics by manual, and these participants lowered the evaluation of manual translations. The evaluation of the proposed system was slightly lower compared to other evaluation items.

6.2 Discussion

We found from participants comments that the trial listening, lyric candidates, and automatic note division features were desirable for many users. In fact, the proposed system received better comparative evaluations than the manual translation task. It was easier to translate lyrics using the proposed system even for users who can read musical scores.

On the other hand, several problems were found. The completed song lyrics were dependent on the lyric candidates. We would like to develop a function to stimulate the creativity of users rather than some of the implemented functions.

Users wanted to optimize the correspondence between the notes and lyrics using the proposed system. The current implementation randomly applies one of two algorithms for automatic note editing. We suppose it cause stress of participants due to the irregularity by applying one of the algorithms randomly. This system should have fixed rules for the musical notes editing function.

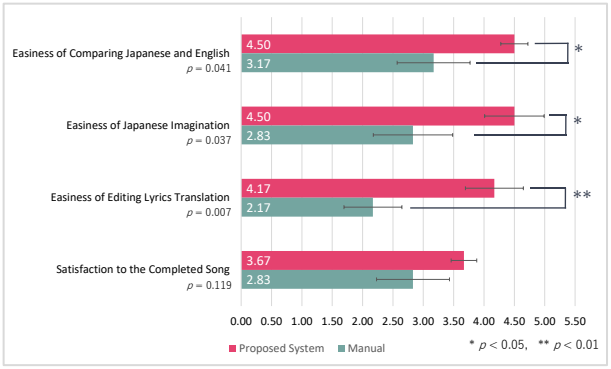

Figure 4. Result of Comparison the Questionnaire Contents.

7 CONCLUSION AND FUTURE WORK

We proposed and implemented an application that supports the translation of English song lyrics into Japanese. We conducted comparative experiments with 12 participants. As a result, we found that lyric translations by the proposed system were evaluated higher than the manual translations. Also, we proposed an algorithm of automatic lyrics translation.

The current implementation of the proposed system randomly applies one of two algorithms to divide notes to match the lyrics and melody. We will improve these algorithms by adopting rules that will allow users to edit notes by simply inputting vowels or stretching bars as lyrics. We would also like to improve the algorithm for automatic temporal generation of the translated lyrics.

REFERENCES

[1] Chihiro Abe and Akinori Ito. 2012. A Japanese Lyrics Writing Support System for Amateur Songwriters. *Signal & Information Processing Association Annual Summit and Conference 2012*.

[2] Marjan Ghazvininejad, Xing Shi, Jay Priyadarshi, and Kevin Knight. 2017. Hafez: an Interactive Poetry Generation System. *Proceedings of ACL 2017, System Demonstrations*.

[3] Hugo Gonçalo Oliveira. 2015. Tra-la-Lyrics 2.0: Automatic Generation of Song Lyrics on a Semantic Domain. *Journal of Artificial General Intelligence* 6, 1: 87–110.

[4] Naoyuki Matsuda. 2013. Introductions for A Comparative Study : Translations of “Orfeo ed Euridice” by Mori Ogai and Kondo Sakufu (in Japanese). *Komazawa Journal of Japanese Culture* 7: 165–192.

[5] Keiichiro Oura, Ayami Mase, Tomohiko Yamada, Satoru Muto, Yoshihiko Nankaku, and Keiichi Tokuda. 2010. Recent development of the HMM-based singing voice synthesis system-Sinsy. *Seventh ISCA Workshop on Speech Synthesis*: 211–216.

[6] Aniruddh D. Patel and Joseph R. Daniele. 2003. An empirical comparison of rhythm in language and music. *Cognition* 87, 1: B35–B45.

[7] Makiko Sadakata, Peter Desain, John R Iversen, Henkjan Honing, and Aniruddh D Patel. 2003. An analysis of rhythm in Japanese and English popular music. *Proceedings of the Annual meeting of the Japanese Society for Music Perception and Cognition*.

[8] M. Utiyama, M., Takahashi. 2003. English-japanese translation alignment data. http://www2.nict.go.jp/astrec-att/member/mutiyama/align/index.html.

[9] Kento Watanabe, Yuichiro Matsubayasi, Kentaro Inui, Tomoyasu Nakano, Satoru Fukayama, and Masataka Goto. 2017. LyriSys: An Interactive Support System for Writing Lyrics Based on Topic Transition. *Proceedings of the 22nd International Conference on Intelligent User Interfaces*: 559–563.

[10] Kazuhide Yamamoto, Satoshi Ikeda, and Kazuteru Ohashi. 2005. Transoforming a Sentence End into News Headline Style. *Natural Language Processing* 12: 85–111.

[11] 2017. Japanese wikipedia corpus. . Retrieved from http://dumps.wikimedia.org/jawiki/latest/jawikilatest-pages-articles.xml.bz2

[12] 2018. Microsoft translator API. Retrieved from https://www.microsoft.com/en-us/translator/.