

# Design of a Japanese Machine Translation System Based on Speech Recognition Technology

Xiaoying Li

lqlixiaoying@126.com

Shandong Vocational College of Science and Technology, No. 6388 West Ring Road, Weifang City, Shandong Province, China

**Abstract.** In order to enhance the accuracy of Japanese machine translation, this study proposes the design of a Japanese machine translation system based on speech recognition technology. The hardware components of the system include a speech recognition hardware selection unit and a Japanese semantic translation hardware selection unit. The software components consist of a Japanese automatic speech recognition module, a Japanese speech-to-semantic conversion module, and a Japanese automatic translation module. Through the design and development of the aforementioned hardware units and software modules, this research has achieved optimized performance of the Japanese machine translation system. The research results demonstrate that the system yields larger values for BLEU and NIST parameters, indicating higher accuracy in Japanese machine translation. The findings strongly indicate that the designed system provides superior Japanese machine translation performance and is suitable for extensive promotion and use.

**Keywords:** Speech recognition, Japanese semantics, Japanese translation, System design

## 1 Introduction

In recent years, with the rapid development of the Chinese economy and increasing trade and tourism exchanges between China and Japan, there has been a growing demand for high-quality Japanese language machine translation systems. Designing a high-performance Japanese machine translation system is one of the pressing challenges in the field of translation today. Niu Libao and Wang Zhenduo (2023) addressed the increasing prevalence of hybrid online and offline teaching in universities and the lag in the development of online examination systems. They designed and developed a Japanese online examination system that is user-friendly, reliable, and practical for students to master Japanese knowledge and engage in fragmented learning. This system aligns with the trend of "Internet+" education [1]. Bu Xue'er (2022) conducted a comparative analysis of various mainstream speech recognition tools available in the market. They ultimately chose "Xunfei Tingjian" as an example and explored the role of speech recognition tools in assisting interpretation [2]. Fu Xiaoxia (2022) identified several characteristics of technical Japanese translation, including the prevalence of loanwords, complex technical vocabulary, and the use of complex sentence structures. In their translation practice, they discovered certain patterns and techniques to address these challenges [3]. Geng Yaohui (2020) employed the theory of "communicative translation" as a guiding framework for their work. They categorized the challenges faced during the

translation process into three main types: vocabulary translation, punctuation handling, and inter-sentence cohesion [4]. Zhan Zhan (2021) focused on speech recognition technology and machine translation as their research subjects to achieve offline speech translation [5]. Ru Kuang (2014) proposed a novel method for automatically extracting translation pairs of named entities from monolingual corpora, considering the characteristics of both Chinese and Japanese languages [6]. It is evident that there is relatively limited research specifically related to Japanese language speech translation systems. Existing language recognition and translation systems are not well-suited for Japanese, leading to suboptimal machine translation results. These limitations hinder effective communication and exchanges between China and Japan, particularly in trade and tourism. Therefore, there is a need to explore and develop a Japanese language machine translation system based on speech recognition. It is hoped that by addressing the intricacies of machine translation technology application, the effectiveness of Japanese language recognition and translation can be improved, fostering stronger communication and interactions between China and Japan, ultimately contributing to the development of both nations.

## **2 Hardware Design of the Japanese Speech Machine Translation System**

The hardware design consists of two units: the machine vision hardware selection unit and the speech-to-semantic translation hardware unit. The specific design process is as follows.

### **2.1 Hardware Selection Unit for the Speech Recognition Module**

The principle of speech recognition technology involves the application of hardware to capture the target signal. Speech recognition hardware primarily includes audio sensors and sound acquisition cards [7]. The selected sensor configuration features a 2/3-inch chip and multiple interface types to meet various hardware connection requirements.

The sound acquisition card is a crucial hardware device for capturing and storing information in speech recognition [8]. The AM9513 chip is selected as the information acquisition card for speech recognition, and its structure is depicted in Figure 1.



**Fig. 1.** Hardware structure of acquisition card

The sound collection card includes digital input, output components, clock components and other components. The maximum frequency of the collection is 20 MHz, which can meet the requirements of Japanese voice recognition and machine translation.

## 2.2 Hardware Unit Design for Speech-to-Semantic Translation

The hardware for Japanese semantic translation consists of a microprocessor chip and peripheral circuits [9]. The STM32F103VET6 processor is selected as the microprocessor chip. Before performing Japanese semantic translation, it is necessary to amplify the collected audio signal to ensure accuracy in recognition. Therefore, the peripheral circuit is designed as an audio signal amplification circuit, as illustrated in Figure 2.

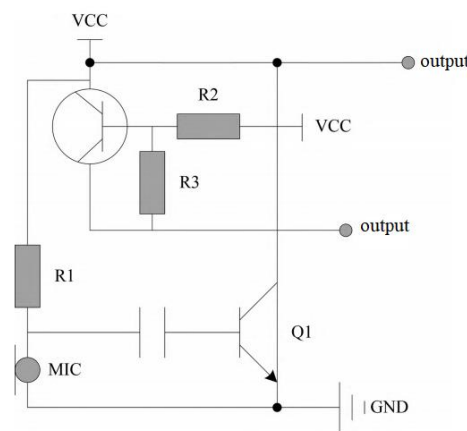


Fig. 2. Schematic diagram of peripheral signal amplification circuit

## 3 Japanese Automatic Recognition and Translation System Software Design

The primary purpose of system software design is to process signals using software in order to extract crucial information from the target. The design of the translation system software modules includes the Japanese speech recognition module, the Japanese speech-to-semantic conversion module, and the Japanese automatic translation module [10]. The specific design process is as follows.

### 3.1 Japanese Speech Recognition Module

Japanese speech recognition employs the Dynamic Time Warping (DTW) algorithm. In the training phase, this algorithm divides the Japanese speech signal into multiple templates. During the recognition phase, the incoming Japanese speech to be recognized is compared with these templates to achieve automatic recognition of Japanese speech. The DTW algorithm is effective in handling template matching problems where pronunciation durations vary, and it boasts advantages such as fast analysis speed and minimal data processing, making it one of the most widely applied algorithms in the field of speech recognition in today's language technology.

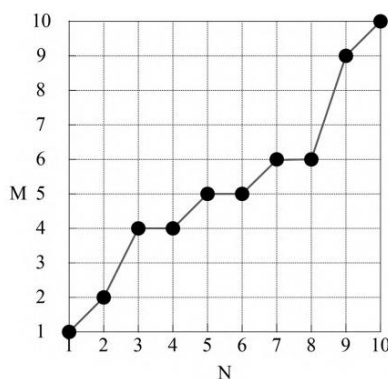
Set the Japanese voice template to:

$R=\{R(1),R(2),\dots,R(m),\dots,R(M)\}$ .  $m$  and  $M$  represent the starting and ending frames of Japanese speech respectively, and  $R(m)$  represent the characteristic vector of Japanese speech in frame  $m$ . Suppose the test template is  $T=\{T(1),T(2),\dots,T(n),\dots,T(N)\}$ , where  $n$  and  $N$  represent the starting and ending frames to be recognized for Japanese speech acquisition, respectively.

The similarity between Japanese speech and template to be recognized is expressed by the distance between the two. The formula is:

$$D[T,R]=\min\sum_{i=1}^N d(T(i),R(i)) \quad (1)$$

If  $N=M$ , the similarity degree can be calculated directly by formula (1). However, the length of Japanese speech to be recognized is usually different from that of template, so it is necessary to align  $T(n)$  with  $R(m)$  by linear expansion method to ensure the correct recognition of Japanese speech. Therefore, the dynamic time regulation algorithm search path is shown in Figure 3.



**Fig. 3.** Dynamic time regulation algorithm search path

As shown in Figure 3, the horizontal and vertical coordinates represent the frame numbers of the Japanese speech and template to be identified. Because the order of speech acquisition is fixed, the algorithm search path starts from the lower left corner and ends in the upper right corner. Figure 3 shows the result of automatic Japanese speech recognition for the voice frame corresponding to the best path, and stores it in text format.

### 3.2 Japanese Speech-to-Semantic Conversion Module Design

Building upon the Japanese speech recognition results obtained, the Japanese speech-to-semantic conversion hardware is employed to extract Japanese semantic information. This module aligns and processes the audio information with semantic information, thereby obtaining a semantic representation of the perceived Japanese speech to prepare for subsequent Japanese automatic translation.

The specific steps for designing the Japanese speech-to-semantic conversion software module are as follows:

Step One: Generate a Japanese word embedding table based on random initialization.

Step Two: Train the word embedding table using information from speech recognition to acquire Japanese semantic features.

Step Three: If the dimensions of the text and semantic features do not match, employ a linear fully connected network to appropriately transform the dimensions, ensuring that the dimensions of the text and semantic features are the same.

Step Four: Compute the dot product between the text and semantic features, and normalize them to obtain an attention matrix.

Step Five: Utilizing the Japanese semantic features obtained in Step Two and the attention matrix obtained in Step Four, integrate operations to obtain a semantic representation of the perceived Japanese text for automatic recognition.

In order to simplify the representation process of Japanese semantic transformation, step 3 text and semantic feature dimension transformation are simplified as follows:

$$\begin{cases} Attention(Q,V) = Soft\ max\left(\frac{Q^T}{\sqrt{d_k}}\right)V \\ MultiHead(Q,V) = Concat(h_1, h_2, \dots, h_n)W^o \\ h_i = Attention(QW_1^o, VW_1^o) \end{cases} \quad (2)$$

In Formula (2), Attention(Q,V) represents the speech-semantic feature dimension conversion function. Q and V represent speech and semantic representation respectively;  $d_k$  represent dimension values after unification; MultiHead(Q,V) represent dimension mean;  $h_i$  represent length values of speech and semantic representation;  $W^o$  represent auxiliary parameter matrix. Through the above process, the Japanese speech semantic conversion module is designed to provide accurate information support for subsequent Japanese machine translation.

### 3.3 Japanese Machine Translation Module Design

Based on the above-mentioned results of Japanese semantic conversion, Japanese phonetic features are extracted, and Japanese machine translation is realized based on relative conditional entropy of words. In the process of translation from Japanese to Chinese, both have cultural characteristics of Chinese characters, so phonetic similar information is taken as one of the important characteristics. The formula for speech similarity probability is:

$$P(J,C) = \frac{\sum p(J_i, C_i)}{n} \quad (3)$$

In Formula (3), J and C represent the phonetic features of Japanese phrases and Chinese phrases, P(J, C) represent the phonetic similarity probability between J and C, and n represent

the number of phrases corresponding words. Based on the calculation result  $P(J,C)$  of formula(3), the relative conditional entropy of words is calculated. The expression is:

$$S(K) = \sum d(K, C_i) = \sum P(C_i|K) \log_{10} \frac{P(C_i, K)}{(P(C_i) + \eta)P(K)} \quad (4)$$

In Formula (4),  $S(K)$  represents the relative conditional entropy of words corresponding to the central word  $K$ ;  $d$  represents the European distance;  $C_i|K$  represents the posterior probability distribution of co-occurrence words in the context of known central words. The correspondence between Japanese phrases and Chinese phrases is calculated. In the process of translation, "one-to-many" or "many-to-one" phenomena are likely to occur. At this time, the relative conditional entropy of words needs to be added to the translation probability calculation.

$$p\left(\frac{j_i^l}{c_1^m}\right) = \sum_A \text{align}(j_1^l, a_1^m | c_1^m) \quad (5)$$

In Formula(5),  $p(i/c)$  denotes the probability of translation;  $A$  denotes the correspondence between Japanese and Chinese phrases;  $a_i$  denotes the alignment;  $i$  denotes the range of values  $[0,1]$ ; and  $l$  and  $m$  denotes the total number of Japanese and Chinese phrases respectively. Look for an alignment method to maximize the result  $p(j/c)$  of formula (5). In this case, the alignment method  $a_1$  is the result of Japanese machine translation.

## 4 Experiment and result analysis

Other language intelligent recognition translation APP is selected as the the comparison system, as shown in Table 1.

**Table 1.** Experimental environment configuration

Name	Configuration
Hardware environment	Intel Core TM 2 Duo CPU T6600 2. 5GHz
translator	SRILM 1. 4. 5
software platform	Linux,ubuntu·12. 04·beta2·desktop·i386
programming language	C++
Identification module open source tool	CMU Sphinx
Conversion module open source tool	DeepSpeech2
Translation module open source tool	SRILM 1. 7. 1

### 4.1 Experimental preparation stage

This research experiment preparation stage mainly undertakes the task of pretreatment of experimental data. In the experiment, 10 groups of Japanese were randomly selected from a Japanese database as experimental data. Due to the influence of lip radiation, the power of Japanese speech signal is inconsistent, and the resolution of speech signal is low, which affects the accuracy of experimental results. Set the experimental data - Japanese voice signal sampling value is  $x(n)$ , the pre-weighted voice signal sampling value expression is:

$$y(n) = x(n) - \alpha x(n-1) \quad (6)$$

In Formula (6),  $y(n)$  represents the pre-emphasized voice signal sampling value;  $\alpha$  represents the pre-emphasized coefficient, the value range is (0.9, 1.0). In addition, rectangular windows are used to add windows to ensure the power stability of speech signals. Rectangular window expressions are:

$$\omega(n) = \begin{cases} 1 & (0 \leq n \leq N-1) \\ 0 & (\text{other}) \end{cases} \quad (7)$$

## 4.2 Experimental results

### 4.2.1 Analysis of BLEU parameters

BLEU parameter calculation formula is:

$$BLEU = BP \left[ \exp \left( \sum_{n=1}^N \omega(n) \log p_n \right) \right] \quad (8)$$

In Formula (8), BP represents the length penalty factor;  $p_n$  represents the cumulative result of the number of occurrences of  $n$ -membered word strings. The BLEU parameter is the key index of BLEU evaluation method. The range of values is [0,1], 0 for worst and 1 for best, which means that the larger the BLEU parameter value, the better the effect of Japanese speech machine translation. The BLEU parameters obtained by experiment are shown in Figure 4.

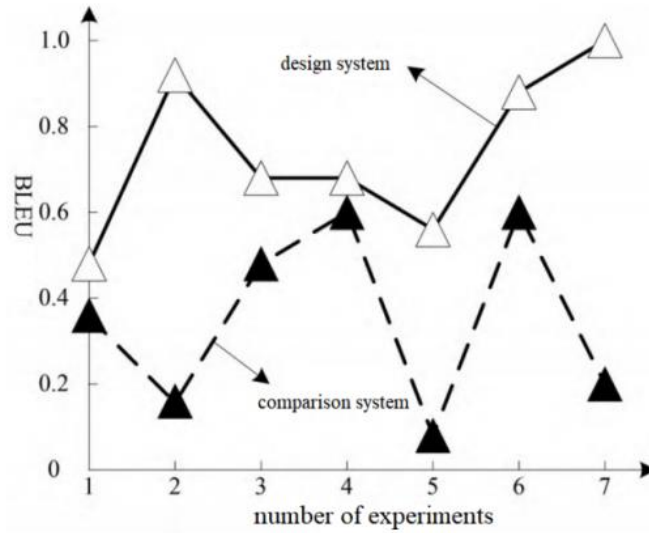


Fig. 4. BLUE Parameters

As shown in Figure 4, compared with the the comparison system, the BLEU parameters obtained by the design system are larger, and the maximum value can reach 1.0.

#### 4.2.2 Analysis of NIST parameters

NIST parameter calculation formula is:

$$NIST = \frac{\sum_{n=1}^N Infow(n)}{\sum_{n=1}^N Infow(1)} \exp \left\{ \beta \log_2 \left[ \min \left( \frac{L_{SYS}}{L_{RES}}, 1 \right) \right] \right\} \quad (9)$$

In Formula (9), Infow(n) represents the information weight corresponding to the n-meta word string,  $\beta$  represents the empirical value,  $L_{SYS}$  represents the length of the translated word string, and  $L_{RES}$  represents the average length of the reference translation.

NIST parameter is the key index of NIST evaluation method. The value range of NIST parameter is [0,1]. With the increase of NIST parameters, the better the effect of automatic Japanese recognition translation.

The NIST parameters obtained by experiment are shown in Figure 5. As shown in Figure 5, the NIST parameters obtained by using the design system are larger than those of the comparison system, and the maximum value can reach 0.9.

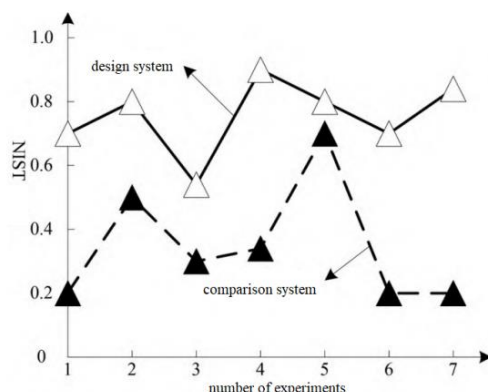


Fig. 5. NIST parameters

#### 4.2.3 Analysis of Japanese Speech Machine Translation Accuracy Rate

The accuracy of Japanese speech machine translation is shown in Table 2.

Table 2. Japanese Automatic Recognition Translation Accuracy Table

Experimental number	design system	the comparison system
1	89.45%	56.23%
2	90.12%	61.45%
3	85.43%	70.12%
4	92.30%	68.39%
5	88.32%	65.20%
6	76.51%	68.48%
7	90.01%	65.37%



As shown in Table 2, the accuracy of Japanese machine translation obtained by using the design system is higher than that of the the comparison system, and the maximum value is 92.30%.The results show that the BLEU parameters and NIST parameters are larger and the accuracy of Japanese automatic recognition translation is higher than that of the comparison system.

## 5 Conclusions

This study aimed to enhance the accuracy of Japanese machine translation by employing a system design approach based on speech recognition technology. Through the construction of hardware units for speech recognition and Japanese semantic translation, complemented by software components including a Japanese automatic speech recognition module, Japanese speech-to-semantic conversion module, and Japanese automatic translation module, a Japanese machine translation system was successfully developed. The research findings indicate a significant improvement in the accuracy of Japanese machine translation. Specifically, the system achieved higher values in both BLEU and NIST parameters, signifying a substantial enhancement in translation quality. The application of speech recognition technology allowed the system to better comprehend spoken language expressions, thereby providing more precise and natural translation results. This research outcome represents a significant advancement in the field of Japanese machine translation, offering robust support for the promotion and application of Japanese-related applications.

## References

- [1] Niu Li Bao and Wang Zhenduo. Japanese Online Test System Based on Speech Recognition Technology. Information Technology, 2023(08):8-12.
- [2] Rusli A, Shishido M. Zero-Pronoun Annotation Support Tool for the Evaluation of Machine Translation on Conversational Texts[J]. Journal of Natural Language Processing, 2022, 29(2): 493-514.
- [3] Fu Xiaoxia, a Chinese pop artist. Translation report of technical terms and long sentences in Japanese. Shanxi University, 2022.
- [4] Geng Yaohui. Translation Practice Report of Artificial Intelligence and Virtual Reality. Dalian University of Foreign Studies, 2020.
- [5] Zhan Zhan. Offline Speech Translation Technology. Hangzhou University of Electronic Science and Technology, 2021.
- [6] Ru Kuang. A Study on the Acquisition Method and Its Application of Japanese - Chinese Bilingual Naming Entity Pair. Beijing Jiaotong University, 2014
- [7] Xu Bo. Study on standardized oral processing based on conditional random field. Nanjing University of Technology, 2010.
- [8] Stentiford F W M, Steer M G. Machine translation of speech[J]. British Telecom technology journal, 1988, 6(2): 116-122.
- [9] Dhanjal A S, Singh W. An automatic machine translation system for multi-lingual speech to Indian sign language[J]. multimedia Tools and Applications, 2022: 1-39.
- [10] Nakamura S, Markov K, Nakaiwa H, et al. The ATR multilingual speech-to-speech translation system[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2006, 14(2): 365-376.