

Original Paper

# Code-Switching Automatic Speech Recognition for Nursing Record Documentation: System Development and Evaluation

Shih-Yen Hou, MS; Ya-Lun Wu, MS; Kai-Ching Chen, MS; Ting-An Chang, PhD; Yi-Min Hsu, PhD; Su-Jung Chuang, MHA; Ying Chang, MS; Kai-Cheng Hsu, MD, PhD

Artificial Intelligence Center for Medical Diagnosis, China Medical University Hospital, Taichung City, Taiwan

**Corresponding Author:**

Kai-Cheng Hsu, MD, PhD

Artificial Intelligence Center for Medical Diagnosis

China Medical University Hospital

No 2, Yude Rd

North Dist

Taichung City, 40459

Taiwan

Phone: 886 +886911284382 ext 168

Email: [D35842@mail.cmuh.org.tw](mailto:D35842@mail.cmuh.org.tw)

## Abstract

**Background:** Taiwan has insufficient nursing resources due to the high turnover rate of health care providers. Therefore, reducing the heavy workload of these employees is essential. Herein, speech transcription, which has various potential clinical applications, was employed for the documentation of nursing records. The requirement of including only one speaker per transcription facilitated data collection and system development. Moreover, authorization from patients was unnecessary.

**Objective:** The aim of this study was to construct a speech recognition system for nursing records such that health care providers can complete nursing records without typing or with only a few edits.

**Methods:** Nursing records in Taiwan are mainly written in Mandarin, with technical terms and abbreviations presented in both Mandarin and English. Therefore, the training set consisted of English code-switching information. Next, transfer learning (TL) and meta-TL (MTL) methods, which perform favorably in code-switching scenarios, were applied.

**Results:** As of September 2021, the China Medical University Hospital Artificial Intelligence Speech (CMaiSpeech) data set was established by manually annotating approximately 100 hours of recordings from 525 speakers. The word error rate (WER) of the benchmark model of syllable-based TL was 29.54% in code-switching. The WER of the proposed model of syllable-based MTL was 22.20% in code-switching. The test set comprised 17,247 words. Moreover, in a clinical case, the proposed model of syllable-based MTL yielded a WER of 31.06% in code-switching. The clinical test set contained 1159 words.

**Conclusions:** This paper has two main contributions. First, the CMaiSpeech data set—a Mandarin-English corpus—has been established. Health care providers in Taiwan are often compelled to use a mixture of Mandarin and English in nursing records. Second, an automatic speech recognition system for nursing record document conversion was proposed. The proposed system can shorten the work handover time and further reduce the workload of health care providers.

(*JMIR Nursing* 2022;5(1):e37562) doi: [10.2196/37562](https://doi.org/10.2196/37562)

**KEYWORDS**

nursing records; automatic speech recognition; code-switching; transfer learning; meta-transfer learning

## Introduction

**Background**

In Taiwan, more than 80% of health care providers are employed in the acute care system, mainly providing care for inpatients. However, the turnover rate of new recruits in medical centers is considerably high, exceeding 22.5%; in regional hospitals, it

is approximately 29% [1-3]. Turnover rate is defined as the number of people leaving per month divided by the average number of people in the month. According to the equation below, Price [4] thinks that the turnover rate exceeds 50%, which is potentially harmful; Gaukerke [5] considers that the ideal annual turnover rate should be approximately 5% to 10%. On average, 1 in 4 new health care providers leaves because they cannot endure the harsh working conditions in hospitals.

Specifically, the reasons for the high turnover rate are presented as follows. First, the clinical workload is not divided into professional and nonprofessional work; thus, delegating nonprofessional work to others is not possible and, as such, nursing personnel are responsible for various menial tasks. Second, the particularity of three-shift work is highly inconvenient [6,7].

$$\text{Turnover Rate} = \frac{\text{Number of people leaving per month}}{\text{Average number of people in the month}} \times 100\%$$

Health care providers spend most of their time in patient rooms, including isolation rooms, and on administrative tasks, including completing nursing records and charting work and reviewing them [8,9]. Descriptions of medical evaluations and their results are included in nursing records such that the medical care team can fully understand the patient's condition. Incomplete or erroneous nursing records are a serious violation of patients' rights. The causes of incomplete nursing records are inadequate human resources and differences in writing styles, which increase cognitive differences during health care providers' shifts and can cause problems in patient care. According to the American Joint Commission International Accreditation Standards for Hospitals, approximately 65% of medical care problems are due to miscommunication [10]. Therefore, nursing records play a vital role in medical care.

Speech recognition programs, which have seen various advances in recent years, can understand human language; their application to the generation of nursing records can potentially improve work efficiency and reduce the workload of health care providers. Many countries have employed speech recognition technology in medicine [11,12]. The speech-to-text automation of nursing records can lighten the burden of administrative work. Although automatic speech recognition in the medical domain was first reported in the 1980s [13], all subsequent studies up to 1999 tested the transcription of single words as opposed to continuous speech in this context [14]. In recent years, a few studies have been conducted on speech recognition in the medical domain in terms of the word error rate (WER) [15-17].

## Related Work

The connectionist temporal classification (CTC) [18] and the listen, attend, and spell (LAS) [19] frameworks have been used to record physician-patient conversations [20]. The corpus size is 14,000 hours. Data cleaning comprised a major portion of the work. The WERs of the CTC and LAS methods were 20.1% and 18.3%, respectively, but both methods can be used to transcribe conversations between health care providers and patients. For the CTC method, a clean corpus and a corresponding language model are integral. Compared with the CTC method, the LAS framework has a higher tolerance to corpora containing a small amount of error. Because all the contextual information is considered by the LAS method, its accuracy is slightly higher than that of other models; however, the LAS method cannot perform streaming of automatic speech recognition. In addition, the length of the input voice segment has a significant impact on the accuracy of the model.

The learning method proposed in Winata et al [21] was designed to overcome the shortage of code-switching data by generating synthetic code-switching sentences. In that study, a

sequence-to-sequence model was presented, and pointer-generator networks were used to generate code-switching data. The model had two principal characteristics: learning how to combine words from parallel sentences and determining when to switch from one language to another. However, the model was applied only to general conversations and not to conversations in a specific field. The main reason is that there are few conversations in the nursing data set, mainly nursing records, including records of patients' physiological conditions, records of treatment and medication, and records of doctors' orders.

## The Goal of This Study

The shortage of nursing labor in Taiwan has been a concern of Taiwan's Ministry of Health and Welfare in recent years, and this situation has been exacerbated and highlighted by the spread of COVID-19 [22,23]. Health care providers in Taiwan have a heavier workload compared with those in other regions and countries. To mitigate this problem, we established a code-switching automatic speech recognition system enabling health care providers to complete nursing records without typing or editing. We used the China Medical University Hospital Artificial Intelligence Speech (CMaiSpeech) data set for this purpose.

## Methods

### CMaiSpeech Data Set

In the medical field, automatic speech recognition poses several difficulties. First, background noise is often present during recordings. Second, speaking speed varies considerably among individuals. Third, speech may contain hesitations and utterances, such as repeated words and restarted sentences. The most formidable challenge is code-switching. Taiwanese people often use a mixture of Chinese and English in daily conversation; this is often observed in the medical field and in nursing records. Medical terms, including thousands of drug names, are often written in English or represented by English abbreviations, meaning that sentences commonly contain a mix of Mandarin and English words.

Nursing records in Taiwan mostly contain Mandarin-English sentences. To obtain a code-switching corpus for training the automatic speech recognition model, a nursing record corpus was created and used as the Mandarin-English code-switching corpus. The process of the CMaiSpeech data set has been created as follows. First, some code-switching sentences were selected from the nursing records in the China Medical University Hospital (CMUH) database. Next, volunteers were recruited to speak and record these sentences. After the recording process was completed, each sentence was annotated based on the recording context. Finally, the audio data were converted into a 16-kHz pulse-code modulation format. Figure 1 displays examples of Mandarin-English code-switching utterances in clinical nursing records.

Code-switching is a common language phenomenon and refers to a person alternately using more than one language or its variants in a conversation. A generalized definition of code-switching includes language alternation within sentences

(ie, words of different languages are in the sentence) and outside the sentences (ie, sentences of other languages are in the data set). For our cases, most language alternations happened within sentences. The CMaiSpeech data set contains numerous medical

terms in both English and Mandarin. The data set was established by manually annotating approximately 100 hours of recordings from 525 speakers. Most of the speakers were nurses employed at CMUH.

**Figure 1.** Examples of Mandarin-English code-switching utterances in clinical nursing records.

病人 GCS E 四 M 六 V 五 此次因吐血解黑便至急診求治。
(Patient GCS E4M6V5 Went to the emergency department for treatment due to vomiting of blood and Black poops.)
左下肢 AK 術後傷口 hemovac 存 leak 嚴重。
(The hemovac of the left lower extremity after the operation of AK is severely leaked.)
病患因 PT prolong 腹水存 經醫師評估後囑 blood transfusion FFP 四 u 依醫師囑輸血前給予 normal saline 五百 ml ivd stat。
(The patient had ascites due to PT prolong The doctor ordered blood transfusion FFP four u Give normal saline 500 ml ivd stat before blood transfusion according to doctor's order.)

## Ethics Approval

This research was approved by the Institutional Review Board of China Medical University Hospital (CMUH), under approval No. CMUH110-REC2-187.

## National Education Radio Data Set

CMaiSpeech is a code-switching data set. But model training requires a Mandarin corpus with a Taiwanese accent. Therefore, National Education Radio (NER) content is considered a Mandarin data set. NER content from the Formosa Speech in the Wild data set [24] is a Mandarin corpus. For the Mandarin corpus, volume 1 of the NER data set, which contains approximately 100 hours of speech recordings, was used. The data were split into training and evaluation sets by the setup described in the source (ie, volume 1 of the NER data set). Both CMaiSpeech and NER data sets were employed in model training, validation, and testing.

## Transfer Learning

Transfer learning (TL) refers to the transfer of a trained model and parameters to a new model such that the new model does not need to be retrained [25]: this is considered positive transfer. In 2020, Joshi et al [26] applied TL to the recurrent neural network transducer model in the field of speech recognition [27,28]. When the learning of one kind of knowledge promotes the learning of another kind of knowledge, it is called positive transfer. On the contrary, when the learning of one kind of knowledge hinders the learning of another kind of knowledge, it is called negative transfer. For example, when the tasks of the source domain and the target domain are not related, this may lead to negative transfer. In this study, the TL model was regarded as a benchmark model, and the CMaiSpeech and NER data sets were used for two-stage training.

## Meta-Transfer Learning

Meta-TL (MTL), proposed in Winata et al [29], is used to overcome the challenges encountered in code-switching speech recognition. Specifically, this model is used to map audio

recordings to graphemes with relatively few resources. MTL is based on model-agnostic meta-learning [30], with a few modifications for the code-switching tasks. These modifications exchange the model's ability to perform a wide range of tasks for a more favorable ability to initialize code-switching tasks.

## Model Selection

The mainstream automatic speech recognition system can be divided into two models: syllable based and character based. Figure 2 shows the difference between the two models. The character-based model uses the hidden Markov model (HMM) to convert syllables to characters. On the contrary, the syllable-based model has no converter. In Taiwan, nursing records are mainly written in Mandarin, with technical terms and abbreviations being written in both Mandarin and English. Accordingly, we selected a method suitable for nursing handover. The syllable-based method [30] is used in Mandarin, wherein Chinese words are converted into syllables to reduce their complexity. Next, a cascade syllable decoder is used to learn the lingual information to map syllables to graphemes.

Although the most commonly used Chinese transliteration system in Taiwan is Bopomofo [31], we employed simplified pinyin (ie, standard Chinese pinyin without tone markers) to standardize the form of the English and Mandarin words at the character level. Task complexity was reduced through vocabulary minimization. For syllable decoding, the HMM was used along with the Viterbi algorithm. A lexicon containing English words was required to decode syllables produced by the speech transformer model into graphemes.

The training and testing flowchart of the proposed MTL models is shown in Figure 3. The training and testing flowchart of the MTL models is made up of three blocks. First, regarding the data set block, the monolingual data set (ie, NER) and code-switching data set (ie, CMaiSpeech) were jointly learned through meta-training; the corpora of the CMaiSpeech and NER data sets were converted to characters or syllables, a step that can be understood as audio conversion into features; and

character-based and syllable-based models were built. Second, regarding the training block, the models were trained using the MTL method proposed by Guzmán et al [15]; all models can then be further fine-tuned. In addition, for the syllable-based

model, an HMM was trained using the CMaiSpeech data set and a lexicon that maps Mandarin characters to pinyin. Third, regarding the testing block, meta-testing was used.

Figure 2. The difference between syllable-based and character-based models.

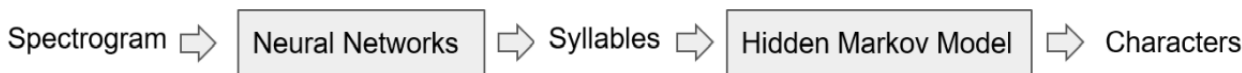
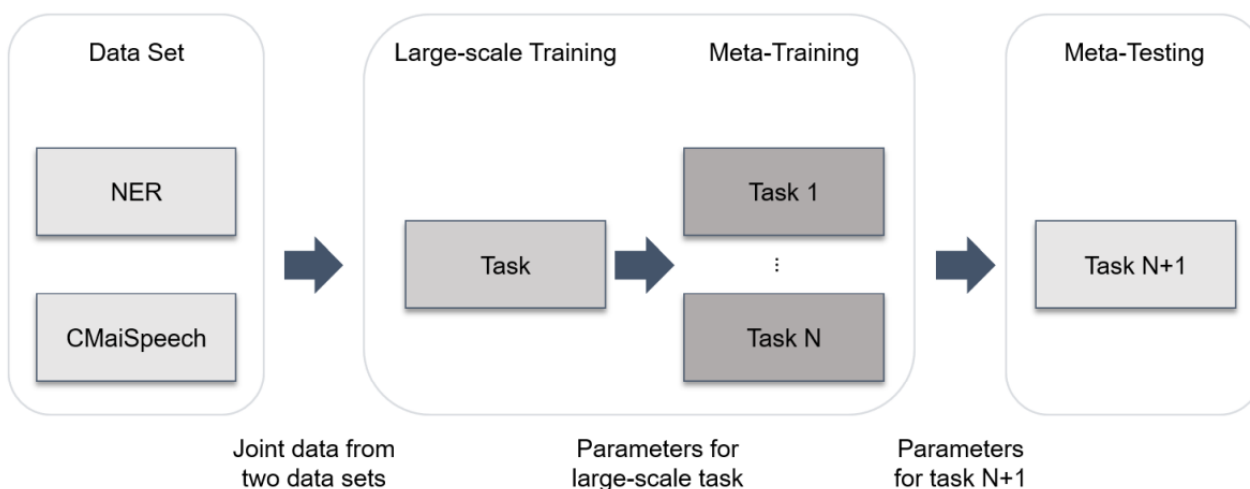


Figure 3. Training and testing flowchart of the meta-transfer learning models. NER: National Education Radio; CMaiSpeech: China Medical University Hospital Artificial Intelligence Speech.



## Results

### Data Set Analysis

The MTL and TL algorithms were evaluated and compared with the CMaiSpeech and NER data sets. The statistics and experimental settings of the data sets are presented in Table 1. Regarding wave statistics, approximately 126 hours of monolingual Mandarin data (ie, NER data set) and approximately 107 hours of code-switching data (ie, CMaiSpeech data set) were employed for training and validation, respectively. Next, the TL and MTL models were assessed using approximately 1.8 hours of code-switching data (ie, CMaiSpeech data set). The CMaiSpeech data set contained 136,196 English words and 787,701 Chinese words, whereas the NER data set contained 682 English words and 1,420,101 Chinese words. On the basis of the proportions of English and Chinese words, the code-switching level of the CMaiSpeech data set was

substantially higher than that of the NER data set. Therefore, the CMaiSpeech data set was determined to be more useful for testing. The testing corpus of the CMaiSpeech data set contained 2427 English words and 14,820 Chinese words. To quantify the code-switching level, the language entropy was used to evaluate the code-switching level. The language entropy of the CMaiSpeech and NER data sets in the training process was 0.6033 and 0.0060, respectively, and the language entropy of the CMaiSpeech data set in the testing process was 0.5857. The probability of switching (ie, the integration index [I-index]) [30] was used to describe the code-switching frequency. The mean I-indexes of the CMaiSpeech and NER data sets in the training process were 0.2006 and 0.0006, respectively, and the mean I-index of the CMaiSpeech data set in the testing process was 0.1411. The frequency of Chinese and English words in the CMaiSpeech data set was determined (Multimedia Appendix 1).

**Table 1.** Statistics and experimental setting of the CMaiSpeech and NER data sets.

Methods	CMaiSpeech <sup>a</sup> data set		NER <sup>b</sup> data set		CMaiSpeech data set
	Training	Validation	Training	Validation	Testing
<b>Wave statistics</b>					
Duration, hours	105.5334	1.6572	114.2523	12.4390	1.8093
Mean duration, seconds	14.2634	18.9395	21.6751	21.2028	17.5566
Utterances, n	26,636	315	18,976	2112	371
<b>Word statistics, n</b>					
English words	136,196	2342	682	150	2427
Chinese words	787,701	12,115	1,420,101	206,909	14,820
Total words	923,897	14,457	1,420,784	207,059	17,247
<b>Code-switching level</b>					
Language entropy	0.6033	0.6391	0.006	0.0086	0.5861
I-index <sup>c</sup> , mean (SD)	0.2006 (0.1529)	0.1896 (0.16)	0.0006 (0.0052)	0.0011 (0.008)	0.1411 (0.1372)

<sup>a</sup>CMaiSpeech: China Medical University Hospital Artificial Intelligence Speech.

<sup>b</sup>NER: National Education Radio.

<sup>c</sup>I-index: integration index (ie, the probability of switching).

## Model Analysis

Table 2 lists the performance of various automatic speech recognition models on CMaiSpeech test sets in WERs. The benchmark models included the character-based TL model, the syllable-based TL model, and the Microsoft Azure cloud application programming interface (API) [32]. The proposed models were the character-based MTL model and the syllable-based MTL model. The total number of words in the testing set was 17,247. The output results were Chinese characters and English words. The test sets had two characteristics: the sentences code-switched between English and Mandarin and the sentences included many drug names and English abbreviations.

Among the evaluated models, the WERs of the character- and syllable-based TL models were 27.22% and 29.54% WER-code-switching (WER-CS), respectively, whereas that of the Azure cloud API was 59.20% WER-CS. By contrast, the WERs of the proposed character- and syllable-based MTL

models were 24.78% and 22.20% WER-CS, respectively. In most cases, the MTL models yielded higher WER values than did the TL models because the CMaiSpeech data set was smaller. The MTL algorithm is suitable for solving small-sample learning problems because the MTL model was jointly trained using both the CMaiSpeech data set and the NER data set. By contrast, the TL algorithm is usually pretrained on a data set to obtain a more favorable initial model, which is then fine-tuned on the new data set. The new data set must have a considerable amount of data; this is typically not the case in a small-sample learning task. Thus, joint training enables the MTL model to have robust code-switching capabilities.

For the same test sets, the performance of syllable-based automatic speech recognition models without HMM was measured on CMaiSpeech test sets (Table S1 in [Multimedia Appendix 2](#)). The WERs of the syllable-based TL and syllable-based MTL models without HMM were 24.16% and 16.85% WER-CS, respectively. The output of the models was syllables, not Chinese characters and English words.

**Table 2.** Performance of various automatic speech recognition models on CMaiSpeech<sup>a</sup> test sets (code-switching) in terms of the word error rate.

Methods	WER-CS <sup>b</sup> , %	WER-EN <sup>c</sup> , %	WER-ZH <sup>d</sup> , %
Character-based transfer learning	27.22	40.15	26.75
Syllable-based transfer learning	29.54	38.13	29.22
Azure cloud application programming interface	59.20	95.57	54.23
Character-based meta-transfer learning (proposed)	24.78	41.23	24.28
Syllable-based meta-transfer learning (proposed)	22.20	36.13	21.54

<sup>a</sup>CMaiSpeech: China Medical University Hospital Artificial Intelligence Speech.

<sup>b</sup>WER-CS: word error rate-code-switching.

<sup>c</sup>WER-EN: word error rate-English.

<sup>d</sup>WER-ZH: word error rate-Zhōngwén.

## External Verification

Two experiments were conducted for external verification. First, to analyze the performance of the proposed models and the public automatic speech recognition system, based on the experimental settings, the same CMaiSpeech test sets were applied to the automatic speech recognition model of the Azure cloud API. The WER of the Azure cloud API was 59.20% WER-CS (Table 2). This may be attributable to the fact that the CMaiSpeech test sets included domain-specific conversations about nursing care, whereas the Azure cloud API mainly included general conversations.

Second, the proposed models were deployed online in the Medical Information Technology Office of CMUH. The clinical handover data were collected and used as clinical test sets to validate various automatic speech recognition models (Table 3). The total number of words in the clinical test sets was 1159. The WER of the syllable-based TL models was 33.65% WER-CS, whereas that of the Azure cloud API was 31.75% WER-CS. As for the proposed syllable-based MTL model, the WER was 31.06% WER-CS. The WERs of the proposed systems from the online results were not as favorable as those from the offline results, possibly because the clinical test sets contained a high proportion of Mandarin words and a low proportion of English words. Therefore, when output English words were misspelled, their WERs were substantially higher than those of the Chinese words. In the future, we must focus

on increasing the size of the training database to optimize the performance of the proposed automatic speech recognition engine.

Interesting experimental results were found, as seen in Table 3. If the test sets were divided into a long sentence part and a short sentence part, with 15 seconds as the boundary, the length was categorized and the WERs were calculated separately, as shown in Table 4.

Figure 4 shows the relationship between sentence length and the WER. From the experimental results, it can be found that the sentence length has an inflection point at 40 words. When the sentence length is more than 40 words, the sentence length is positively correlated with the WER, and this result is understandable. On the contrary, when the sentence length is less than 40 words, the sentence length is negatively correlated with the WER. There are two main reasons for this. First, uncommon words in short sentences increase the probability of identification errors. Second, there are limited previous occurrences that can be referenced. Both of these reasons may increase the WER value. In addition, the relationship between the code-switching level and the WER was determined (Multimedia Appendix 3). From the experimental results, it can be found that the code-switching level has an inflection point at 0.40. When the code-switching level is more than 0.40, the code-switching level is positively correlated with the WER.

**Table 3.** Performance of the benchmark and proposed models on clinical test sets in the clinical field in terms of the word error rate.

Methods	WER-CS <sup>a</sup> , %	WER-EN <sup>b</sup> , %	WER-ZH <sup>c</sup> , %
Character-based transfer learning	60.40	107.59	56.88
Syllable-based transfer learning	33.65	85.58	30.31
Azure cloud application programming interface	31.75	96.08	27.80
Character-based meta-transfer learning (proposed)	43.57	114.06	38.41
Syllable-based meta-transfer learning (proposed)	31.06	89.52	27.84

<sup>a</sup>WER-CS: word error rate-code-switching.

<sup>b</sup>WER-EN: word error rate-English.

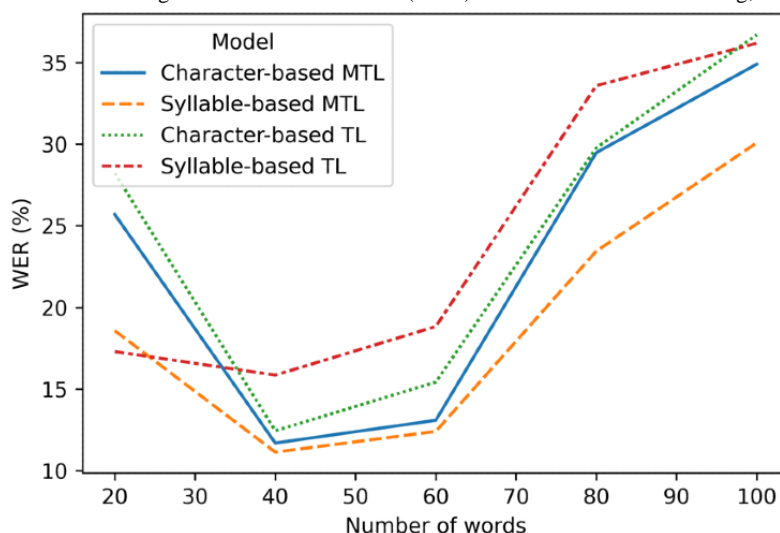
<sup>c</sup>WER-ZH: word error rate-Zhōngwén.

**Table 4.** Performance of the benchmark and proposed models on long and short sentences in terms of the word error rate.

Methods	Long sentences, WER <sup>a</sup> , %	Short sentences, WER, %
Character-based transfer learning	33.58	10.38
Syllable-based transfer learning	35.55	13.63
Character-based meta-transfer learning (proposed)	30.87	8.66
Syllable-based meta-transfer learning (proposed)	27.14	9.11

<sup>a</sup>WER: word error rate.

**Figure 4.** The relationship between sentence length and the word error rate (WER). MTL: meta-transfer learning; TL: transfer learning.



## Discussion

In this study, a Mandarin-English code-switching speech recognition system was developed using a corpus of nursing records to allow such records to be completed without typing or editing. The proposed system can lighten the heavy workloads of medical personnel. The WER of the benchmark syllable-based TL model was 29.54% WER-CS, and that of the proposed syllable-based MTL model was 22.20% WER-CS. The test sets

comprised 17,247 words. In a real clinical case, the WER of the proposed model was 31.06% WER-CS. The clinical test sets contained 1159 words.

Future studies should focus on training with large Mandarin corpora to improve the performance of the proposed code-switching speech recognition engine. We plan to release the proposed corpus of nursing data with the accompanying processing software to the research and development community for the study of clinical language processing.

## Acknowledgments

This research project was conducted using data pulled from the data warehouse of electronic medical records that were provided by China Medical University Hospital. We thank the National Center for High-performance Computing of the National Applied Research Laboratories in Taiwan for providing computational and storage resources. The study was funded by China Medical University and China Medical University Hospital (grants MOST 110-2314-B-039-010-MY2, MOST 111-2321-B-039-005, MOST 111-2622-8-039-001 -IE, MOST 110-2321-B-039-002, and DMR-111-227).

## Authors' Contributions

K-CH and Y-LW contributed to the conception and design of the study. Y-MH, S-JC, and YC collected the data. S-YH designed and developed the system. KCC performed the data analysis. T-AC wrote the manuscript. All authors reviewed and provided feedback for each draft of the manuscript, and all authors read and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Frequency of Chinese and English words in the CMaiSpeech data set. CMaiSpeech: China Medical University Hospital Artificial Intelligence Speech.

[[PNG File , 26 KB-Multimedia Appendix 1](#)]

## Multimedia Appendix 2

Performance of syllable-based automatic speech recognition models without the hidden Markov model (HMM) on the CMaiSpeech test set. CMaiSpeech: China Medical University Hospital Artificial Intelligence Speech.

[[DOCX File , 13 KB-Multimedia Appendix 2](#)]

### Multimedia Appendix 3

The relationship between the code-switching level and the word error rate (WER). I-index: integration index; MTL: meta-transfer learning; TL: transfer learning.

[\[PNG File , 116 KB-Multimedia Appendix 3\]](#)

### References

1. Kao CC. Multi-aspects of nursing manpower in Taiwan. *Cheng Ching Med J* 2011;7(3):41-46. [doi: [10.30156/CCMJ.201107.0005](https://doi.org/10.30156/CCMJ.201107.0005)]
2. Chou S. [Response of Taiwan nursing education to today's nursing shortage] [Article in Chinese]. *Hu Li Za Zhi* 2012 Oct;59(5):24-29. [doi: [10.6224/JN.59.5.24](https://doi.org/10.6224/JN.59.5.24)] [Medline: [23034544](https://pubmed.ncbi.nlm.nih.gov/23034544/)]
3. Lin P, MacLennan S, Hunt N, Cox T. The influences of nursing transformational leadership style on the quality of nurses' working lives in Taiwan: A cross-sectional quantitative study. *BMC Nurs* 2015;14:33 [FREE Full text] [doi: [10.1186/s12912-015-0082-x](https://doi.org/10.1186/s12912-015-0082-x)] [Medline: [25991910](https://pubmed.ncbi.nlm.nih.gov/25991910/)]
4. Price JL. The impact of turnover on the organization. *Work Occup* 2016 Aug 17;16(4):461-473. [doi: [10.1177/0730888489016004005](https://doi.org/10.1177/0730888489016004005)]
5. Gauerke RD. Appraisal as a retention tool. *Superv Nurse* 1977;8(6):34-37. [doi: [10.1097/00006247-197706000-00006](https://doi.org/10.1097/00006247-197706000-00006)]
6. Lin C, Huang C, Kao C, Lu M. [The nursing shortage and nursing retention strategies in Taiwan] [Article in Chinese]. *Hu Li Za Zhi* 2013 Jun;60(3):88-93. [doi: [10.6224/JN.60.3.88](https://doi.org/10.6224/JN.60.3.88)] [Medline: [23729345](https://pubmed.ncbi.nlm.nih.gov/23729345/)]
7. Huang C, Yu C, Yu C. [An Exploration of Working Conditions and Flexible System for Hospital Nurses] [Article in Chinese]. *Hu Li Za Zhi* 2016 Apr;63(2):80-90. [doi: [10.6224/JN.63.2.80](https://doi.org/10.6224/JN.63.2.80)] [Medline: [27026560](https://pubmed.ncbi.nlm.nih.gov/27026560/)]
8. Lee S, Pai H, Yen W. [A study of perceived practice environment among nurses in Taiwan] [Article in Chinese]. *Hu Li Za Zhi* 2008 Aug;55(4):30-40. [Medline: [18668480](https://pubmed.ncbi.nlm.nih.gov/18668480/)]
9. Yeh M, Yu S. Job stress and intention to quit in newly-graduated nurses during the first three months of work in Taiwan. *J Clin Nurs* 2009 Dec;18(24):3450-3460. [doi: [10.1111/j.1365-2702.2009.02941.x](https://doi.org/10.1111/j.1365-2702.2009.02941.x)] [Medline: [20487491](https://pubmed.ncbi.nlm.nih.gov/20487491/)]
10. Aiken LH, Cheung RB, Olds DM. Education policy initiatives to address the nurse shortage in the United States. *Health Aff (Millwood)* 2009;28(4):w646-w656 [FREE Full text] [doi: [10.1377/hlthaff.28.4.w646](https://doi.org/10.1377/hlthaff.28.4.w646)] [Medline: [19525285](https://pubmed.ncbi.nlm.nih.gov/19525285/)]
11. Johnson M, Lapkin S, Long V, Sanchez P, Suominen H, Basilakis J, et al. A systematic review of speech recognition technology in health care. *BMC Med Inform Decis Mak* 2014 Oct 28;14:94 [FREE Full text] [doi: [10.1186/1472-6947-14-94](https://doi.org/10.1186/1472-6947-14-94)] [Medline: [25351845](https://pubmed.ncbi.nlm.nih.gov/25351845/)]
12. Chung SL, Chen YS, Su SF, Ting HW. Preliminary study of deep learning based speech recognition technique for nursing shift handover context. In: Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics. 2019 Presented at: The 2019 IEEE International Conference on Systems, Man and Cybernetics; October 6-9, 2019; Bari, Italy p. 528-533. [doi: [10.1109/smc.2019.8913954](https://doi.org/10.1109/smc.2019.8913954)]
13. Dudley H, Riesz R, Watkins S. A synthetic speaker. *J Franklin Inst* 1939 Jun;227(6):739-764. [doi: [10.1016/s0016-0032\(39\)90816-1](https://doi.org/10.1016/s0016-0032(39)90816-1)]
14. Furui S. Automatic speech recognition and its application to information extraction. In: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics. 1999 Presented at: The 37th Annual Meeting of the Association for Computational Linguistics; June 20-26, 1999; College Park, MD p. 11-20. [doi: [10.3115/1034678.1034680](https://doi.org/10.3115/1034678.1034680)]
15. Guzmán G, Ricard J, Serigos J, Bullock B, Toribio A. Metrics for modeling code-switching across corpora. In: Proceedings of the Conference of the International Speech Communication Association (INTERSPEECH). 2017 Presented at: the Conference of the International Speech Communication Association (INTERSPEECH); August 20-24, 2017; Stockholm, Sweden p. 67-71 URL: [https://www.isca-speech.org/archive/pdfs/interspeech\\_2017/guzman17\\_interspeech.pdf](https://www.isca-speech.org/archive/pdfs/interspeech_2017/guzman17_interspeech.pdf) [doi: [10.21437/interspeech.2017-1429](https://doi.org/10.21437/interspeech.2017-1429)]
16. Klakow D, Peters J. Testing the correlation of word error rate and perplexity. *Speech Commun* 2002 Sep;38(1-2):19-28. [doi: [10.1016/s0167-6393\(01\)00041-3](https://doi.org/10.1016/s0167-6393(01)00041-3)]
17. Wang YY, Acero A, Chelba C. Is word error rate a good indicator for spoken language understanding accuracy. In: Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding. 2003 Presented at: The IEEE Workshop on Automatic Speech Recognition and Understanding; November 30-December 4, 2003; St Thomas, VI p. 577-582. [doi: [10.1109/asru.2003.1318504](https://doi.org/10.1109/asru.2003.1318504)]
18. Graves A, Fernández S, Gomez F, Schmidhuber J. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd International Conference on Machine Learning. 2006 Presented at: The 23rd International Conference on Machine Learning; June 25-29, 2006; Pittsburgh, PA p. 369-376. [doi: [10.1145/1143844.1143891](https://doi.org/10.1145/1143844.1143891)]
19. Chan W, Jaitly N, Le QV, Vinyals O. Listen, attend and spell. ArXiv Preprint posted online on August 20, 2015 [FREE Full text]
20. Chiu CC, Tripathi A, Chou K, Co C, Jaitly N, Jaunzeikare D, et al. Speech recognition for medical conversations. In: Proceedings of the Conference of the International Speech Communication Association (INTERSPEECH). 2018 Presented at: The Conference of the International Speech Communication Association (INTERSPEECH); September 2-6, 2018;



- Hyderabad, India p. 2972-2976 URL: [https://www.isca-speech.org/archive/pdfs/interspeech\\_2018/chiu18\\_interspeech.pdf](https://www.isca-speech.org/archive/pdfs/interspeech_2018/chiu18_interspeech.pdf) [doi: [10.21437/interspeech.2018-40](https://doi.org/10.21437/interspeech.2018-40)]
21. Winata GI, Madotto A, Wu CS, Fung P. Code-switched language models using neural based synthetic data from parallel sentences. In: Proceedings of the 23rd Conference on Computational Natural Language Learning. 2019 Presented at: The 23rd Conference on Computational Natural Language Learning; November 3-4, 2019; Hong Kong, China p. 271-280. [doi: [10.18653/v1/k19-1026](https://doi.org/10.18653/v1/k19-1026)]
  22. Lin C, Lu MS, Huang HY. The psychometric properties and the development of the indicators of quality nursing work environments in Taiwan. *J Nurs Res* 2016 Mar;24(1):9-20. [doi: [10.1097/jnr.000000000000106](https://doi.org/10.1097/jnr.000000000000106)] [Medline: [26670132](https://pubmed.ncbi.nlm.nih.gov/26670132/)]
  23. Lin CF, Huang CI, Yang CM, Lu MS. The relationship between work environment satisfaction and retention intention among nursing administrators in Taiwan. *J Nurs Res* 2019 Oct;27(5):e43 [FREE Full text] [doi: [10.1097/jnr.0000000000000312](https://doi.org/10.1097/jnr.0000000000000312)] [Medline: [31524854](https://pubmed.ncbi.nlm.nih.gov/31524854/)]
  24. Formosa Speech in the Wild. URL: <https://sites.google.com/speech.ntut.edu.tw/fsw/home> [accessed 2022-10-18]
  25. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Proceedings of the 28th Conference on Neural Information Processing Systems. 2014 Presented at: The 28th Conference on Neural Information Processing Systems; December 8-13, 2014; Montréal, QC p. 1-9 URL: <https://proceedings.neurips.cc/paper/2014/file/375c71349b295f2cdca9206f20a06-Paper.pdf>
  26. Joshi V, Zhao R, Mehta RR, Kumar K, Li J. Transfer learning approaches for streaming end-to-end speech recognition system. In: Proceedings of the Conference of the International Speech Communication Association (INTERSPEECH). 2020 Presented at: The Conference of the International Speech Communication Association (INTERSPEECH); October 25-29, 2020; Shanghai, China p. 2152-2156 URL: [https://www.isca-speech.org/archive/pdfs/interspeech\\_2020/joshi20\\_interspeech.pdf](https://www.isca-speech.org/archive/pdfs/interspeech_2020/joshi20_interspeech.pdf) [doi: [10.21437/interspeech.2020-2345](https://doi.org/10.21437/interspeech.2020-2345)]
  27. Graves A. Generating sequences with recurrent neural networks. ArXiv Preprint posted online on June 5, 2014 [FREE Full text]
  28. Graves A. Sequence transduction with recurrent neural networks. ArXiv Preprint posted online on November 14, 2012 [FREE Full text] [doi: [10.48550/arXiv.1211.3711](https://doi.org/10.48550/arXiv.1211.3711)]
  29. Winata GI, Cahyawijaya S, Lin Z, Liu Z, Xu P, Fung P. Meta-transfer learning for code-switched speech recognition. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 Presented at: The 58th Annual Meeting of the Association for Computational Linguistics; July 5-10, 2020; Virtual p. 3770-3776 URL: <https://aclanthology.org/2020.acl-main.348.pdf> [doi: [10.18653/v1/2020.acl-main.348](https://doi.org/10.18653/v1/2020.acl-main.348)]
  30. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning. 2017 Presented at: The 34th International Conference on Machine Learning; August 6-11, 2017; Sydney, Australia p. 1126-1135.
  31. The Manual of the Phonetic Symbols of Mandarin Chinese. Taipei, Taiwan: Ministry of Education; 2000. URL: [https://language.moe.gov.tw/001/Upload/files/site\\_content/M0001/juyin/index.html](https://language.moe.gov.tw/001/Upload/files/site_content/M0001/juyin/index.html) [accessed 2022-10-18]
  32. Copeland M, Soh J, Puca A, Manning M, Gollob D. Microsoft Azure: Planning, Deploying, and Managing Your Data Center in the Cloud. New York, NY: Apress; 2015.

## Abbreviations

- API:** application programming interface  
**CMaiSpeech:** China Medical University Hospital Artificial Intelligence Speech  
**CMUH:** China Medical University Hospital  
**CTC:** connectionist temporal classification  
**HMM:** hidden Markov model  
**I-index:** integration index  
**LAS:** listen, attend, and spell  
**MTL:** meta-transfer learning  
**NER:** National Education Radio  
**TL:** transfer learning  
**WER:** word error rate  
**WER-CS:** word error rate-code-switching

*Edited by E Borycki; submitted 28.02.22; peer-reviewed by M Lee, D Chrimes; comments to author 13.04.22; revised version received 06.09.22; accepted 25.09.22; published 07.12.22*

*Please cite as:*

*Hou SY, Wu YL, Chen KC, Chang TA, Hsu YM, Chuang SJ, Chang Y, Hsu KC*

*Code-Switching Automatic Speech Recognition for Nursing Record Documentation: System Development and Evaluation*

*JMIR Nursing 2022;5(1):e37562*

*URL: <https://nursing.jmir.org/2022/1/e37562>*

*doi: [10.2196/37562](https://doi.org/10.2196/37562)*

*PMID: [36476781](https://pubmed.ncbi.nlm.nih.gov/36476781/)*

©Shih-Yen Hou, Ya-Lun Wu, Kai-Ching Chen, Ting-An Chang, Yi-Min Hsu, Su-Jung Chuang, Ying Chang, Kai-Cheng Hsu. Originally published in JMIR Nursing (<https://nursing.jmir.org>), 07.12.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Nursing, is properly cited. The complete bibliographic information, a link to the original publication on <https://nursing.jmir.org/>, as well as this copyright and license information must be included.