2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP) | 978-1-7281-6994-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISCSLP49672.2021.9362075

Rnn-transducer With Language Bias For End-to-end Mandarin-English Code-switching Speech Recognition

Shuai Zhang[1,2], *Jiangyan Yi*[2], *Zhengkun Tian*[1,2], *Jianhua Tao*[1,2,3], *Ye Bai*[1,2]

[1]School of Artificial Intelligence, University of Chinese Academy of Sciences, China
[2]NLPR, Institute of Automation, Chinese Academy of Sciences, China
[3]CAS Center for Excellence in Brain Science and Intelligence Technology, China

{shuai.zhang, jiangyan.yi, zhengkun.tian, jhtao, ye.bai}@nlpr.ia.ac.cn

Abstract

Recently, language identity information has been utilized to improve the performance of end-to-end code-switching (CS) speech recognition task. However, previous work use an additional language identification (LID) model as an auxiliary module, which increases computation cost. In this work, we propose an improved recurrent neural network transducer (RNN-T) model with language bias to alleviate the problem. We use the language identities to bias the model to predict the CS points. This promotes the model to learn the language identity information directly from transcriptions, and no additional LID model is needed. We evaluate the approach on a Mandarin-English CS corpus SEAME. Compared to our RNN-T baseline, the RNN-T with language bias can achieve 16.2% and 12.9% relative mixed error reduction on two test sets, respectively.

Index Terms: Code-switching, speech recognition, end-to-end, recurrent neural network transducer, language bias

1. Introduction

Code-switching (CS) speech is defined as the alternation of languages in an utterance. It is a pervasive communicative phenomenon in multilingual communities. Therefore, developing a CS speech recognition (CSSR) system is of great interest.

However, the CS scenario presents challenges to recognition system [1, 2]. Some attempts based on deep neural networks-hidden Markov model (DNN-HMM) framework have been made to alleviate these problems [3, 4]. The methods usually contain components including acoustic, language, and lexicon models that are trained with different objective separately, which would lead to sub-optimal performance. And the design of complicated lexicon including different languages would consume lots of human efforts.

Therefore, end-to-end framework for CSSR has received increasing attention recently [5, 6, 7, 8, 9, 10]. These methods combine acoustic, language, and lexicon models into a single model with joint training. However, the lack of CS training data limits the performance of these methods. To address the problem, language identity information is utilized to improve the performance of recognition [5, 6, 7]. They are usually based on connectionist temporal classification (CTC) [11, 12] or attention-based encoder-decoder models [13, 14] or the combination of both. However, previous work usually use an additional language identification (LID) model as an auxiliary module, which increases the computation cost. Recurrent neural network transducer (RNN-T) is an improved model based on CTC, it augments with a prediction network, which is explicitly conditioned on the previous outputs [15, 16, 17, 18, 19]. And the RNN-T model trained with large speech corpus performs competitively compared to the state-of-art model in some tasks [12]. The prediction network can combine the LID information in a natural way without increasing the computation cost. To the best of our knowledge, RNN-T has not been used for CSSR task.

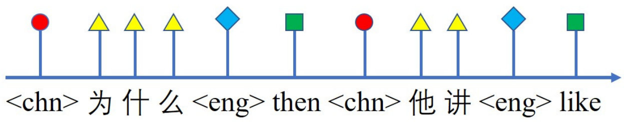

Figure 1: *Code-switching distribution diagram.*

In this paper, we propose an improved RNN-T model with language bias to alleviate the problem. The model is trained to predict language IDs as well as the subwords. To ensure the model can learn CS information, we add language IDs in the CS point of transcription, as illustrated in Fig. 1. In the figure, we use the arrangements of different geometric icons to represent the CS distribution. Compared with normal text, the tagged data can bias the RNN-T to predict language IDs in CS points. So our method can model the CS distribution directly without additional LID model. Then we constrain the input word embedding with its corresponding language ID, which is beneficial for model to learn the language identity information from transcription. In the inference process, the predicted language IDs are used to adjust the output posteriors. The experiment results on SEAME (South East Asia Mandarin English) corpus show that our proposed method outperforms the RNN-T baseline (without language bias) obviously. Overall, our best model achieves 16.2% and 12.9% relative error reduction on two test sets, respectively. To our best knowledge, this is the first attempt of using the RNN-T model with language bias as an end-to-end CSSR strategy.

The rest of the paper is organized as follows. In Section 2, we review RNN-T model. In Section 3, we describe the intuition of the proposed model. In Section 4, we present the experimental setups, and in Section 5, we report and discuss the experiment results in detail. Finally, we conclude the paper in Section 6.

2. Review of RNN-T

Although CTC has been applied successfully in the speech recognition task, it assumes that outputs at each step are independent of the previous predictions [11]. RNN-T is an improved model based on CTC. It augments with a prediction network, which is explicitly conditioned on the previous outputs [16], as illustrated in Fig. 2.

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T)$ be the acoustic input sequence, where T is the frame number of sequence. Let $\mathbf{Y} = (y_1, y_2, ..., y_U)$ be the corresponding sequence of output targets (without language IDs) over the RNN-T output space $\mathcal{Y}$, and $\mathcal{Y}^*$ be the set of all possible sequence over $\mathcal{Y}$. For ASR, the input sequence is much longer than output targets, i.e., $T > U$. Because the frame-level alignments of the target label are unknown, RNN-T augments the output set with an additional symbol, refers to as the *blank* symbol, denoted as ϕ, i.e., $\bar{\mathcal{Y}} \in \mathcal{Y} \cup \{\phi\}$. We denote $\hat{\mathbf{Y}} \in \bar{\mathcal{Y}}^*$ as an alignment, which are equivalent to $(y_1, y_2, y_3) \in \mathcal{Y}^*$ after operation $\mathcal{B}$, such as $\hat{\mathbf{Y}} = (y_1, \phi, y_2, \phi, \phi, y_3) \in \bar{\mathcal{Y}}^*$. Given the input sequence $\mathbf{X}$, RNN-T models the conditional probability $P(\mathbf{Y} \in \mathcal{Y}^*|\mathbf{X})$ by marginalizing over all possible alignments:

$$P(\mathbf{Y} \in \mathcal{Y}^*|\mathbf{X}) = \sum_{\hat{\mathbf{Y}} \in \mathcal{B}^{-1}(\mathbf{Y})} P(\hat{\mathbf{Y}}|\mathbf{X}) \tag{1}$$

where $\mathcal{B}$ is the function that removes consecutive identical symbols and then removing any blank from a given alignment in $\bar{\mathcal{Y}}^*$. And $\mathcal{B}^{-1}(\mathbf{Y})$ is all possible alignments with $\{\phi\}$.

An RNN-T model consists of three different networks as illustrated in Fig. 2. (a) Encoder network (referred to as transcription network) maps the acoustic features into higher level representation $\mathbf{h}_t^{enc} = f^{enc}(\{\mathbf{x}_\tau\}_{1\leq\tau\leq t})$. (b) Prediction network produces output vector $\mathbf{p}_u = f^{pred}(\{y_v\}_{1\leq v\leq u-1})$ based on the previous non-blank input label. (c) Joint network computes logits by combining the outputs of the previous two networks $z_{t,u} = f^{joint}(\mathbf{h}_t^{enc}, \mathbf{p}_u)$. These logits are then passed to a softmax layer to define a probability distribution. The model can be trained by maximizing the log-likelihood of $P(\mathbf{Y} \in \mathcal{Y}^*|\mathbf{X})$.

3. RNN-T with Language Bias

3.1. Output symbols set with language IDs

For this task, we augment the output symbols set with language IDs $< chn >$ and $< eng >$ as shown in Fig. 3, i.e., $\hat{\mathcal{Y}} \in \bar{\mathcal{Y}} \cup \{< chn >, < eng >\}$, $\hat{\mathcal{Y}}$ is the final output symbols set. The intuition behind it is that the CS in the transcript may obey a certain probability distribution, and this distribution can be learned by neural network.

3.2. Properties of RNN-T

The properties of RNN-T is key for the problem. It can predict rich set of target symbols such as speaker role and "end-of-word" symbol, which are not related to the input feature directly [20, 21]. So the language IDs can also be treated as the output symbols. What's more, RNN-T can seamlessly integrate the acoustic and linguistic information. The prediction network of it can be viewed as an RNN language model which predicts the current label given history labels [16]. So it is effective in incorporating LID into the language model. In general, predicting language IDs only from text data is difficult. However, the joint training mechanism of RNN-T allows it to combine the language and acoustic information to model the CS distribution. Furthermore, the tagged text can bias the RNN-T to predict language IDs which indicates CS points, yet the model trained with normal text can not do this. That is why we choose RNN-T to build the end-to-end CSSR system.

3.3. Word Embedding Constraint

To promote the model to learn CS distribution more efficient, we concatenate a short vector to all the English word embedding and the English tag $< eng >$ embedding, another different vector for Mandarin, as shown at the bottom of Fig. 3. This enhances the dependence of word embedding to its corresponding language ID. In the training process, RNN-T model can learn the distinction information between the two languages easily. The experiment results show that the word embedding constraint is an effective technology.

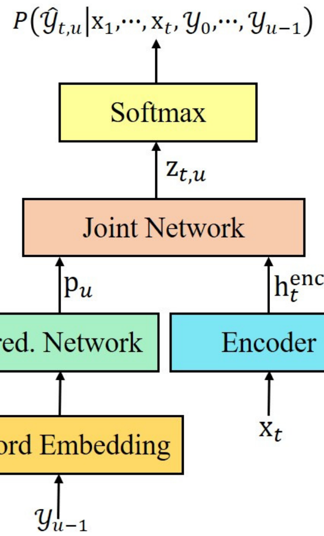

Figure 2: *Basic RNN-T model.*

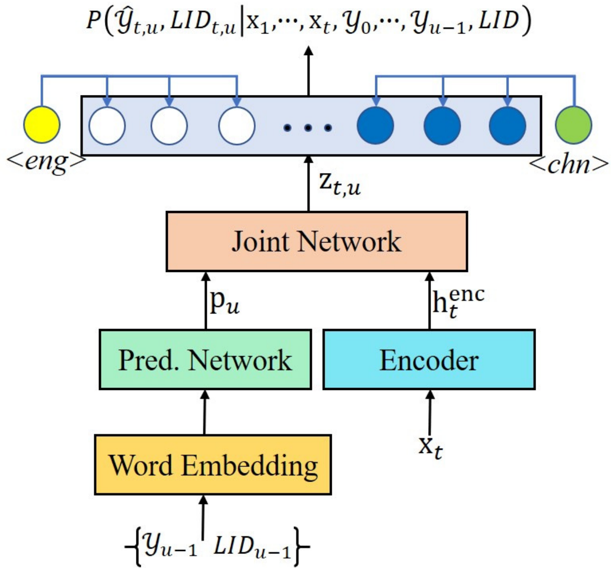

Figure 3: *RNN-T with language bias.*

3.4. Language IDs Re-weighted Decode

In the decode process, we use the predicted language ID to adjust the output posteriors, as shown at the head of Fig. 4. λ is the re-weight scale.This can bias the model to predict a certain language words more likely in the next-step decode. Overall, our proposed method can handle the speech recognition and LID simultaneously in a simple way, and without increasing additional cost. This study provides new insights into the CS information of text data and its application in end-to-end CSSR system. As a

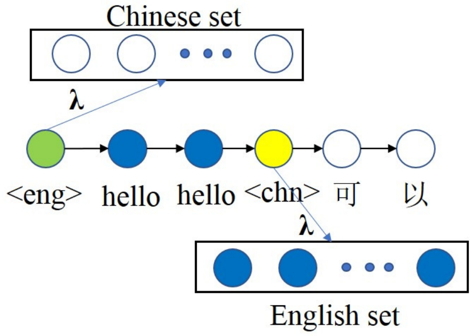

Figure 4: *Language IDs re-weighted decode process.*

final note, the training and inference algorithms of the proposed model are similar to the standard RNN-T model.

4. Experiments Setups

In this section, we introduce in detail the data set, the RNN-T model structure, the modeling unit, and evaluation metrics used in our experiment. We strive to clarify the details of our experiment.

4.1. Overall Data Set Information

We conduct experiments on SEAME, a spontaneous conversational bilingual speech corpus [22]. Most of the utterances contain both Mandarin and English uttered by interviews and conversations in Malaysia and Singapore areas. In this data set, Singapore speakers tend to speak more English in their utterances, and Malaysian speakers tend to speak more Chinese. Specifically, the rate of speakers who speak code-switching utterances with Mandarin rate ranged from 10% to 90% is about 90%, while the rate of speakers who have utterances with Mandarin rates no more than 10% is about 2%, and 5% speakers have utterances over 90% being Mandarin.

4.2. Test Data Set

We use the standard data partitioning rule of previous work which consists of three parts: $train$, $eval_{sge}$ and $eval_{man}$ (see Table 1) [4]. Each evaluation data set is randomly selected from 10 gender balanced speakers. However, $eval_{sge}$ is biased to Southeast Asian accent English speech and $eval_{man}$ is biased to Mandarin speech. The biased data sets are more conducive to show the effectiveness of each proposed methods on each individual languages. The training data set has 101.13h and includes 134 speakers. And the code-switching utterances rate is 68%, which facilitates the learning model to deal with code-switching scene.

Building an end-to-end model requires lots of training data, to improve the reliability of the model, we apply speech speed perturbation to augment speech data [23]. This method has been proven to be a very effective method to increase the amount of data. By manipulation, we get 3 times the data, with the speed rate of 0.9, 1, and 1.1 of the original speech respectively. We use the augmented data to build our RNN-T system.

4.3. RNN-T System

We construct the RNN-T baseline system as described in Section 3. The encoder network of RNN-T model consists of 4 layers of 512 long short-term memory (LSTM). The prediction network is 2 layers with 512 LSTM units. And the joint network consists of single feed-forward layer of 512 units with tanh activate function.

The input acoustic features of encoder network are 80-dimensional log Mel-filterbank with 25ms windowing and 10ms frame shift. Mean and normalization is applied to the features. And the input words embedding of prediction network is in 512 dimensions continuous numerical vector space. During training, the ADAM algorithm is used as the optimization method, we set the initial learning rate as 0.001 and decrease it linearly when there is no improvement on the validation set. To reduce the over-fitting problem, the dropout rate is set to 0.2 throughout all the experiments. In the inference process, the beam-search algorithm [15] with beam size 35 is used to decode the model. All the RNN-T models are trained from scratch using PyTorch.

Table 1: *Data Statistics of SEAME [4]*

Set	Speakers	Hours	Duration Ratio (%)		
			Man	En	CS
$train$	134	101.13	16	16	68
$eval_{man}$	10	7.49	14	7	79
$eval_{sge}$	10	3.93	6	41	53

4.4. Wordpieces

For Mandarin-English CSSR task, it is a natural way to construct output units by using characters. However, there are several thousands of Chinese characters and 26 English letters. Meanwhile, the acoustic counterpart of Chinese character is much longer than English letter. So, the character modeling unit will result in significant discrepancy problem between the two languages. To balance the problem, we adopt BPE subword [24] as the English modeling units. The targets of our RNN-T baseline system contains 3090 English wordpieces and 3643 Chinese characters. The BPE subword units can not only increase the duration of English modeling units but also maintain a balance unit number of two languages.

4.5. Evaluation Metrics

In this paper, we use mixed error rate (MER) to evaluate the experiment results of our methods. The MER is defined as the combination of word error rate (WER) for English and character error rate (CER) for Mandarin. This metrics can balance the Mandarin and English error rates better compared to the WER or CER.

5. Results and Analysis

In this part, we show the experimental results with different settings and conduct an in-depth analysis of these results. Experimental results prove the effectiveness of our method compared to the baseline system.

5.1. Baseline of Standard RNN-T Model

Table 2 reports the main experiment results of our baseline system. Because the amount of data is small, the data augmentation is effective for our baseline system. Because the data augmentation technology can significantly reduce the MER of end-to-end model, we conduct all the following experiments based on augmented training data. To further improve the performance of the system, we use N-gram and neural language model to re-score the N-best (N=35) results. The language models used in this paper are trained with training transcription by Kaldi toolkit [25]. Furthermore, We can also observe that all the experiment results in $eval_{sge}$ is much worse than $eval_{man}$. This is probably that the accent English in data $eval_{sge}$ is more difficult for the recognition system. Bilinguals usually have serious accent problem, which poses challenge to CSSR approaches.

5.2. Results of RNN-T with Language Bias

Table 3 reports the main experiment results of RNN-T with language bias. In order to fairly compare the results of proposed methods with baseline, we remove all the language IDs in the decoded transcription. We can find that the performance of RNN-T model trained (without word embedding constraint) with tagged transcription is much better than the RNN-T baseline. It achieves 9.3% and 7.6% relative MER reduction on two test sets respectively. This shows that the tagged text can improve the modeling ability of RNN-T for the CSSR problem. It is the main factor that causes the MER reduction in our experiments. Furthermore, word embedding constraint can also improve the performance of the system though not significant. Overall, our proposed methods yields improved results without increasing additional training or inference burden.

Table 2: *The results of RNN-T baseline.*

Model	Data Aug	LM	MER(%)	
			$eval_{man}$	$eval_{sge}$
RNN-T	No	No	37.9	49.7
	Yes	No	33.3	44.9
	Yes	4-gram	32.4	42.8
	Yes	RNN-LM	31.8	42.1

Table 3: *The MER of RNN-T with language bias method.*

Model	LM	MER(%)	
		$eval_{man}$	$eval_{sge}$
RNN-T	No	33.3	44.9
+ LID	No	30.2	41.5
++ Emb constraint	No	29.5	40.3

5.3. Effect of Language IDs Re-weighted Decode

We then evaluate the system performance by adjusting the weights of next-step predictions in decode process. Table 4 shows the result of RNN-T model with language IDs re-weighted in decoding. The result shows that this technique can further improve the performance of the model compared to the ordinary decoding process. This suggests that the predicted language IDs can effectively guide the model decoding.

Because the model assigns language IDs to the recognized words directly, the language IDs error rate is hard to compute. It is difficult for us to measure it in a simple and direct way. However, this result may imply that the prediction accuracy of our method is high enough to guide decoding. Meanwhile, We also find that the re-weighted method is more effective on the $eval_{man}$ than $eval_{sge}$. This could be caused by higher language IDs prediction accuracy in $eval_{man}$. After different experiments, we set the re-weight scale $\lambda = 0.2$ in the following experiments.

Table 4: *The result of language IDs re-weighted decode method in inference.*

Model	λ	MER(%)	
		$eval_{man}$	$eval_{sge}$
RNN-T + LID + Emb	No	29.5	40.3
	0.2	28.9	39.7

Table 5: *The MER of language model re-score.*

Model	LM	MER(%)	
		$eval_{man}$	$eval_{sge}$
RNN + LID + Emb	No	28.9	39.7
	4-gram	28.7	39.3
	RNN-LM	28.1	38.9

5.4. Results of Language Model Re-score

Table 5 shows the MER results of the N-best (N=35) re-scoring with N-gram and neural language models. The language models are both trained with the tagged training transcription. We see that the language re-scoring can further improve the performance of models. And the performance of language model re-scoring is better than the model without re-scoring slightly. It reveals that the prediction network of RNN-T still has room to be further optimization.

Finally, compared with the RNN-T baseline with data augment, the proposed method can achieve 16.2% and 12.9% relative MER reduction. And the final model with neural language model re-scoring achieve 11.6% and 7.6% relative MER reduction compared with baseline with re-score. For both scenarios, our RNN-T methods can achieve better performance than baselines.

6. Conclusions and Future Work

In this work we develop an improved RNN-T model with language bias for end-to-end Mandarin-English CSSR task. Our method can handle the speech recognition and LID simultaneously, no additional LID system is needed. Finally, compared with the RNN-T baseline with data augment, the proposed method can achieve 16.2% and 12.9% relative MER reduction. It yields consistent improved results of MER without increasing training or inference cost.

In the future we intend to expand our work in two directions. For the first direction, we plan to pre-train the prediction network of RNN-T model using large text corpus, and then fine-tune the RNN-T model with labeled speech data by frozen the prediction network. This can effectively utilize large amount of Mandarin-English text data which without corresponding speech. Another way is to explore the effectiveness of the attention-based model language bias for end-to-end Mandarin-English CSSR task.

7. Acknowledgements

This work is supported by the National Key Research & Development Plan of China (No.2018YFB1005003), the National Natural Science Foundation of China (NSFC) (No.61831022, No.61901473, No.61771472, No.61773379) and Inria-CAS Joint Research Project (No.173211KYSB20190049).

8. References

[1] Ö. Çetinoğlu, S. Schulz, and N. T. Vu, "Challenges of computational processing of code-switching," *arXiv preprint arXiv:1610.02213*, 2016.

[2] S. Sitaram, K. R. Chandu, S. K. Rallabandi, and A. W. Black, "A survey of code-switched speech and language processing," *arXiv preprint arXiv:1904.00784*, 2019.

[3] Y. Li, P. Fung, P. Xu, and Y. Liu, "Asymmetric acoustic modeling of mixed language speech," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 5004–5007.

[4] P. Guo, H. Xu, L. Xie, and E. S. Chng, "Study of semi-supervised approaches to improving english-mandarin code-switching speech recognition," *arXiv preprint arXiv:1806.06200*, 2018.

[5] Z. Zeng, Y. Khassanov, V. T. Pham, H. Xu, E. S. Chng, and H. Li, "On the end-to-end solution to mandarin-english code-switching speech recognition," *arXiv preprint arXiv:1811.00241*, 2018.

[6] C. Shan, C. Weng, G. Wang, D. Su, and L. Xie, "Investigating end-to-end speech recognition for mandarin-english code-switching," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019.

[7] K. Li, J. Li, G. Ye, R. Zhao, and Y. Gong, "Towards code-switching asr for end-to-end ctc models," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6076–6080.

[8] H. Seki, S. Watanabe, T. Hori, J. L. Roux, and J. R. Hershey, "An end-to-end language-tracking speech recognizer for mixed-language speech," in *Icassp IEEE International Conference on Acoustics*, 2018.

[9] B. Li, Y. Zhang, T. Sainath, Y. Wu, and W. Chan, "Bytes are all you need: End-to-end multilingual speech recognition and synthesis with bytes," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5621–5625.

[10] S. Kim and M. L. Seltzer, "Towards language-universal end-to-end speech recognition," in *ICASSP 2018 - 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[11] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 369–376.

[12] R. Prabhavalkar, K. Rao, T. N. Sainath, L. Bo, and N. Jaitly, "A comparison of sequence-to-sequence models for speech recognition," in *Interspeech 2017*, 2017.

[13] W. Chan, N. Jaitly, Q. Le, and O. Vinyals, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 4960–4964.

[14] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, and Y. Bengio, "End-to-end attention-based large vocabulary speech recognition," in *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2016, pp. 4945–4949.

[15] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 6645–6649.

[16] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *International conference on machine learning*, 2014, pp. 1764–1772.

[17] K. Rao, H. Sak, and R. Prabhavalkar, "Exploring architectures, data and units for streaming end-to-end speech recognition with rnn-transducer," in *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2017, pp. 193–199.

[18] Z. Tian, J. Yi, J. Tao, Y. Bai, and Z. Wen, "Self-attention transducers for end-to-end speech recognition," *arXiv preprint arXiv:1909.13037*, 2019.

[19] J. Li, R. Zhao, H. Hu, and Y. Gong, "Improving rnn transducer modeling for end-to-end speech recognition," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2019, pp. 114–121.

[20] Y. He, R. Prabhavalkar, K. Rao, W. Li, A. Bakhtin, and I. McGraw, "Streaming small-footprint keyword spotting using sequence-to-sequence models," in *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 2017, pp. 474–481.

[21] L. E. Shafey, H. Soltau, and I. Shafran, "Joint speech recognition and speaker diarization via sequence transduction," *arXiv preprint arXiv:1907.05337*, 2019.

[22] D.-C. Lyu, T.-P. Tan, E.-S. Chng, and H. Li, "Mandarin-english code-switching speech corpus in south-east asia: Seame," *Language Resources and Evaluation*, vol. 49, no. 3, pp. 581–600, 2015.

[23] T. Ko, V. Peddinti, D. Povey, M. L. Seltzer, and S. Khudanpur, "A study on data augmentation of reverberant speech for robust speech recognition," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 5220–5224.

[24] C.-C. Chiu, T. N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R. J. Weiss, K. Rao, E. Gonina *et al.*, "State-of-the-art speech recognition with sequence-to-sequence models," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4774–4778.

[25] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," in *IEEE 2011 workshop on automatic speech recognition and understanding*, no. CONF. IEEE Signal Processing Society, 2011.