Report on the Vietnamese ASR task

Nguyen Van Huy
Thai Nguyen University of Technology

***Abstract*— This paper describes the ASR task for VLSP-2018, and an overview of participated systems such as feature extraction, acoustic model, language mode, and decoding process.**

Keywords—VLSP 2018, ASR, VAIS, Viettel.

I. DEFINITION

In the ASR task, participants were asked to transcribe automatically Vietnamese audio files into the spoken word sequences. The committee provided the test set only, while the training data for the acoustic and language models was developed by the teams themselves. The test set was delivered on March 8, 2018 via email, and then each team had two days to recognize it. The final result had to submit by 16-PM on March 10, 2018. The Word Error Rate (WER) was measured with references which were human transcripts of the audio files.

II. TEST DATA

The test set was composed of 796 continuous wav files of news speech for a total duration of two hours, without any information on the sentence segmentation. The speech was recorded in a non-noisy environment, and available in three dialects: Northern, Southern and Central with respectively proportion of 50%, 40% and 10%.

III. SUBMISSION

We received two submissions which were from VAIS and Viettel-CSC.

IV. OVERVIEW OF PARTICIPATED ASR SYSTEMS

	VAIS	**Viettel-CSC**
Resources		
Acoustic Data	1200h, 3 dialects	500h, mostly is Northern
Text data	...	900MB, online newspapers
Words in Lexicon	6,5k	11k (6k Vietnamese + 5k foreign)
System		
Feature	MFCC + Pitch	MFCC + Pitch + Bottleneck
Acoustic model (AM)	7 layers DNN	Combination DNN model including two sub-models (TDD and BLSTM)
Language model (LM)	N-gram	N-gram, RNN-LM
Decoding process	- 1st decoding: AM + 4-gram LM - Re-scoring: 5-gram	- 1st decoding: TDNN + 4-gram, BLSTM + 4 gram - Rescoring: RNN-LM - Combining: Using two 1st-decoding outputs - LM Adaptation: Topic adaptation - 2nd decoding: Re-decoding with Adapted-LM

V. EVALUATION RESULT

Scores in WER and Sentence Error Rate (SER) for the submissions of VAIS and Viettel-CSC are shown in Table 1.

TABLE I. ASR EVALUATION RESULTS FOR VLSP-2018

Team	**WER**	**SER**
VAIS	**6.29%**	75.50%
Viettel-CSC	7.40%	75.38%

VI. CONCLUSIONS

This year is the first time VLSP starts the ASR task. There were six registrations, but only two submissions. The best team in WER is VAIS, but in SER is Viettel-CSC.