

---

# FastCorrect: Fast Error Correction with Edit Alignment for Automatic Speech Recognition

## Supplementary Material

---

Yichong Leng<sup>1</sup>, Xu Tan<sup>2</sup>, Linchen Zhu<sup>3</sup>, Jin Xu<sup>4</sup>, Renqian Luo<sup>1</sup>, Linqun Liu<sup>3</sup>  
Tao Qin<sup>2</sup>, Xiang-Yang Li<sup>1</sup>, Edward Lin<sup>3</sup>, Tie-Yan Liu<sup>2</sup>

<sup>1</sup>University of Science and Technology of China, <sup>2</sup>Microsoft Research Asia

<sup>3</sup>Microsoft Azure Speech, <sup>4</sup>Tsinghua University

<sup>1</sup>{lyc123go,lrq}@mail.ustc.edu.cn,xiangyangli@ustc.edu.cn

<sup>2</sup>{xuta,taoqin,tyliu}@microsoft.com

<sup>3</sup>{linczhu,linqun,edlin}@microsoft.com

<sup>4</sup>j-xu18@mails.tsinghua.edu.cn

## A Appendix

### A.1 Source code of FastCorrect

The source code of FastCorrect can be found in *source\_code.zip*. In zip file, the link of FastCorrect model and autoregressive model that are fine-tuned on the AISHELL-1 dataset can be found.

Note that we only include the test scripts, the recognition results of AISHELL-1 ASR model and the model link in the source code. We will release 1) the ASR model with its Espnet training configurations, 2) pretrained FastCorrect model on 400M crawled data, 3) full training scripts of FastCorrect, and 4) fine-tuned FastCorrect model on AISHELL-1 once our paper is accepted.

More detail about reproducing the results of Table 1 can be found in README of *source\_code.zip*.

### A.2 Homophone dictionary

The homophone dictionary can be easily created by the following steps.

First, we can collect text data (e.g., from Wikipedia) and get a vocabulary.

Second, we can obtain the pronunciation (i.e., the phoneme sequence) of every word in the vocabulary by open-source grapheme-to-phoneme tools (e.g., g2pM<sup>1</sup> package for Mandarin or g2p<sup>2</sup> package for English).

Third, we can calculate the similarity between the phoneme sequences of any two words in the vocabulary. Two words with a similarity higher than threshold are regarded as homophones. In our implementation, the similarity is the negative of ratio of edit distance of two phoneme sequences and average length of two phoneme sequences.

---

<sup>1</sup><https://github.com/kakaobrain/g2pM>

<sup>2</sup><https://github.com/Kyubyong/g2p>