|  |  |  |  |
| --- | --- | --- | --- |
| Title | Non-autoregressive end-to-end neural modeling for automatic pronunciation error detection | | |
| Author(s) Name | Md Anwar Hussen Wadud, Mohammed Alatiyyah, MF Mridha | | |
| Contact Email(s) | firoz.mridha@aiub.edu | | |
| Published Journal Name | Applied Sciences | | |
| Type of Publication | Journal | | |
| Volume | 13 | Issue | 1 |
| Publisher | MDPI | | |
| Publication Date | 2022/12/22 | | |
| ISSN | 2227-7390 | | |
| DOI | 10.3390/app13010109 | | |
| URL | https://www.mdpi.com/2076-3417/13/1/109 | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| A crucial element of computer-assisted pronunciation training systems (CAPT) is the mispronunciation detection and diagnostic (MDD) technique. The provided transcriptions can act as a teacher when evaluating the pronunciation quality of finite speech. The preceding texts have been entirely employed by conventional approaches, such as forced alignment and extended recognition networks, for model development or for enhancing system performance. The incorporation of earlier texts into model training has recently been attempted using end-to-end (E2E)-based approaches, and preliminary results indicate efficacy. Attention-based end-to-end models have shown lower speech recognition performance because multi-pass left-to-right forward computation constrains their practical applicability in beam search. In addition, end-to-end neural approaches are typically data-hungry, and a lack of non-native training data will frequently impair their effectiveness in MDD. To solve this problem, we provide a unique MDD technique that uses non-autoregressive (NAR) end-to-end neural models to greatly reduce estimation time while maintaining accuracy levels similar to traditional E2E neural models. In contrast, NAR models can generate parallel token sequences by accepting parallel inputs instead of left-to-right forward computation. To further enhance the effectiveness of MDD, we develop and construct a pronunciation model superimposed on our approach’s NAR end-to-end models. To test the effectiveness of our strategy against some of the best end-to-end models, we use publicly accessible L2-ARCTIC and SpeechOcean English datasets for training and testing purposes where the proposed model shows the best results than other existing models. | |