

於是用於評分的文本皆有經過文字轉拼音，以便系統合成，由表 5 可知，文本進行中文斷詞後，MOS 有些微的增加，另外，進行評估數字正規化的文本，為突顯正規化的效果，文本皆選用含有阿拉伯數字的中文句子，正規化後 MOS 分數由 4.02 提高到了 4.45，分數大幅的提升了，由此可知，數字正規化對於文本的重要。在英文結果的部份，選用在英文句中含有連續大寫的文本，用以評估處理頭字語的效果，然而在加入頭字語處理後，MOS 分數由 3.69 增加至 3.99，由結果可知透過前處理能提升合成之品質。

5 結論

我們建立的中英文語碼轉換語音合成系統，其有相當不錯的表現，透過中、英文的資料前處理大幅提升語音的品質，尤其是中文的數字正規化與英文的頭字語處理，分別由 4.02 上升至 4.45，及 3.69 至 3.99，不過整體系統依舊有進步的空間，因此，未來也將持續改進語碼轉換中，中英文的語音流暢度，以及以建立一個可分離語者資訊，單純學習文本資訊的編碼器為目標，無需再使用生成模型生成的資料集進行訓練，依然可合成多語言語碼轉換的句子。

References

- Ye Bai, Jianhua Tao, Jiangyan Yi, Zhengqi Wen, and Cunhang Fan. 2018. *Clmad: A chinese language model adaptation dataset*. In *2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 275–279.
- Anmol Gulati, James Qin, Chung-Cheng Chiu, Niki Parmar, Yu Zhang, Jiahui Yu, Wei Han, Shibo Wang, Zhengdong Zhang, Yonghui Wu, et al. 2020. Conformer: Convolution-augmented transformer for speech recognition. *arXiv preprint arXiv:2005.08100*.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Jungil Kong, Jaehyeon Kim, and Jaekyoung Bae. 2020. Hifi-gan: Generative adversarial networks for efficient and high fidelity speech synthesis. *Advances in Neural Information Processing Systems*, 33:17022–17033.
- Wei-Yun Ma and Keh-Jiann Chen. 2005. Design of ckip chinese word segmentation system. *Chinese and Oriental Languages Information Processing Society*, 14(3):235–249.
- Yi Ren, Chenxu Hu, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao, and Tie-Yan Liu. 2020. Fast-speech 2: Fast and high-quality end-to-end text to speech. *arXiv preprint arXiv:2006.04558*.
- Yao Shi, Hui Bu, Xin Xu, Shaoji Zhang, and Ming Li. 2020. Aishell-3: A multi-speaker mandarin tts corpus and the baselines. *arXiv preprint arXiv:2010.11567*.
- J Sun. 2012. Jieba chinese word segmentation tool.
- Yih-Wen Wang. 2021. Integrating hidden speaker and style information to multi-lingual and code-switching speech synthesis. <https://hdl.handle.net/11296/du785x>.
- Shinji Watanabe, Takaaki Hori, Shigeki Karita, Tomoki Hayashi, Jiro Nishitoba, Yuya Unno, Nelson Enrique Yalta Soplín, Jahn Heymann, Matthew Wiesner, Nanxin Chen, et al. 2018. Espnet: End-to-end speech processing toolkit. *arXiv preprint arXiv:1804.00015*.
- Junichi Yamagishi, Christophe Veaux, Kirsten MacDonald, et al. 2019. Cstr vctk corpus: English multi-speaker corpus for cstr voice cloning toolkit (version 0.92).