Disentangled representation online resource

1. J.-C. Chou et al., "Multi-target voice conversion without parallel data by adversarially learning disentangled audio representations," *arXiv preprint arXiv:1804.02812*, 2018.
   - Abstract：利用編解碼器學習 linguistic content，然後可以將語音轉換成不同的 speakers。
   - Paper link：https://arxiv.org/abs/1804.02812
   - Demo link： https://jjery2243542.github.io/voice_conversion_demo/

2. (9/28報告的論文) K. Qian et al., "Global rhythm style transfer without text transcriptions," *arXiv preprint arXiv:2106.08519*, 2021.
   - Abstract：利用編解碼器學習非平行文本之間的節奏訊息
   - Paper link： https://arxiv.org/abs/2106.08519
   - Demo link： https://auspicious3000.github.io/AutoPST-Demo/

3. K. Qian et al., "Unsupervised speech decomposition via triple information bottleneck," *arXiv preprint arXiv:2004.11284*, 2020.
   - Abstract：提出了一種叫做 SpeechSplit 的演算法，可以分別對音色、音高和節奏進行 style transfer。
   - Paper link： https://arxiv.org/abs/2004.11284
   - Demo link： https://auspicious3000.github.io/SpeechSplit-Demo/

4. V. John et al., "Disentangled representation learning for non-parallel text style transfer," *arXiv preprint arXiv:1808.04339*, 2018.
   - Abstract：在非平行文本中學習並解開語音中的 content 和 style，並與先前最先進的方法相比得到更好的 transfer accuracy、content preservation 和 language fluency。
   - Paper link： https://arxiv.org/abs/1808.04339
   - Demo link：沒有提供

Disentangled representation online resource

5. A. Nagrani et al., "Disentangled speech embeddings using cross-modal self-supervision," *arXiv preprint arXiv:2002.08742*, 2020.
   - Abstract：透過影片中人臉與音訊之間的關係，學習 linguistic content 和 speaker identity。
   - Paper link： https://arxiv.org/abs/2002.08742
   - Demo link：沒有提供

6. W. Hsu et al., "Disentangling correlated speaker and noise for speech synthesis via data augmentation and adversarial factorization," *IEEE ICASSP,* 2019, pp. 5901-5905
   - Abstract：因為不同的錄音設備會收錄到不同的背景噪音，因此要學習訊號中的 speaker identity 和 background noise 是具有挑戰的，作者提出三種解決的方法，並證明提出的方法可以學習到其中隱藏的表徵，而且可以合成出 clean speech。
   - Paper link： https://ieeexplore.ieee.org/document/8683561
   - Demo link：https://google.github.io/tacotron/publications/adv_tts/
   - 在 GitHub 上有表示使用些微不同的訓練模型，結果效能不如作者，demo 連結已失效

7. A. Polyak, et al., “Speech resynthesis from discrete disentangled self-supervised representations,” *arXiv preprint arXiv:2104.00355*, Jul. 2021.
   - Abstract： 利用 self-supervised 解開了 speech content、prosodic information 和 speaker identity，並完成了超輕量級的編解碼器。
   - paper link：https://arxiv.org/abs/2104.00355
   - Demo link：https://speechbot.github.io/resynthesis/index.html

8. D. Wang et al., “VQMIVC: vector quantization and mutual Information-based unsupervised speech representation disentanglement for one-shot voice conversion” *arXiv preprint arXiv:2106.10132*, 2021.
   - Abstract ：This paper proposes a speech representation disentanglement framework for one-shot/any-to-any voice conversion, which performs conversion across arbitrary speakers with only a single target-speaker utterance for reference.
   - paper link： https://arxiv.org/abs/2106.10132
   - Demo link：https://replicate.ai/wendison/vqmivc