DLHLP HW2 Voice Conversion Report

組長 github id:

組員：

**HW2-1 (Auto-Encoder) (2.5%)**

1. 請以 Auto-Encoder 之方法實做 Voice conversion。jo如果同學不想重新刻一個 auto-encoder，可以試著利用[這個repo](https://github.com/BogiHsu/Voice-Conversion)的部分程式碼，達到實現出 auto-encoder。如果你是修改助教提供的 repo，請在 report 當中敘述你是如何更改原本程式碼，建議可以附上修改部分的截圖以利助教批閱；同時，果餓未有更動原本模型參數也請一併列出。如果你的 auto-encoder是自己刻的，那也請你簡單敘述你的實作方法，並附上對應程式碼的截圖。(1%)
2. 在訓練完成後，試著將助教要求轉換的音檔轉成 source speaker 和 target speaker 的 interpolation，也就是在 testing 的時候，除了將指定的音檔轉成 p1 和 p2 的聲音之外，請嘗試轉成p1 和 p2 interpolation 的聲音。並比較分析 interpolated 的聲音和 p1 以及 p2 的關係。你可以從聲音頻率的高低、口音、語調等面向進行觀察。只要有合理分析助教就會給分。請同時將這題的音檔放在github 的 hw2-1資料夾中，檔名格式請參考投影片。(1.5%)

**HW2-2 (GAN) (2.5%)**

1. 請使用助教在投影片中提到的連結，進行 voice conversion。請描述在這個程式碼中，語者資訊是如何被嵌入模型中的？請問這樣的方式有什麼優缺點？有沒有其他的作法可以將 speaker information 放入generator 裡呢？(1%)
2. 請描述你如何將原本的程式碼改成訓練兩個語者的 voice conversion 程式。 (0.5%)
3. 請問這個程式碼中，input acoustic feature 以及 generator output 分別是什麼呢？ (1%) Hint: 請研究一下 preprocess 時做了哪些事情。

**HW2-3 (1) 和 (2) 擇一回答 (4%)**

1. 請自己找一個不是 StarGAN-VC，也不是 HW2-1 的 model，實際 train 看看。請詳細描述 model 架構， training objective，訓練時是否需要 paired data 等等。(4%) Hint: [useful link](https://zerospeech.com/2019/)
2. 想辦法 improve HW2-1或是 HW2-2 的 model (或是改一些有趣的東西)。Hint:各位可以想想看 speaker embedding 有沒有什麼其他方式？如果今天我在 testing 的時候想要讓他有 unseen speaker 也可以成功轉過去的話，用什麼 embedding會比較好？(hint: d-vector, i-vector) 又或者要怎麼把這個 speaker embedding 餵進 model 裡面呢？有什麼[不同的方法](https://arxiv.org/abs/1802.05637)？