

Statistical Parametric Synthesis Implement

Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

Outline

Action item

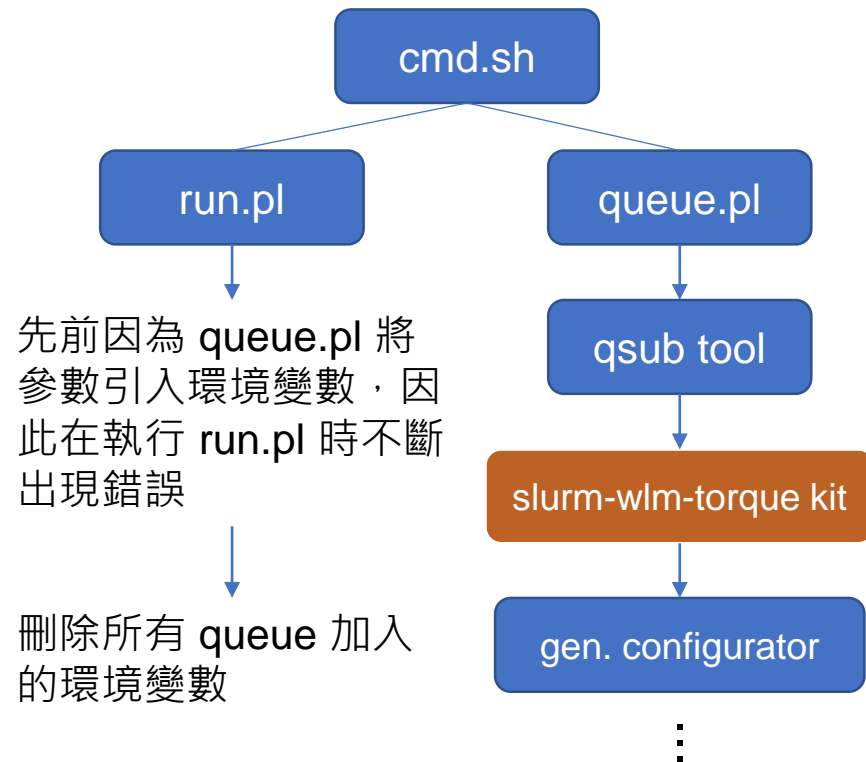
- Introduce the building blocks of conventional TTS

Status report

- 預期目標：執行完程式、合成出音檔後，再看程式中所執行的步驟。(目前仍在執行程式階段)
- 執行步驟
 - 安裝 concatenative speech synthesis TTS 工具 (IdiapTTS)、環境以及各套件
 - 以 IdiapTTS 工具為主進行實驗腳本，執行腳本前須下載預訓練模型與下載 LJSpeech 資料庫
 - 使用 source cmd.sh 連結腳本與 IdiapTTS 工具 (問題1)
 - 實驗步驟可分為三大項
 - **特徵生成**
 - 創建資料庫 (問題2)
 - 創建強制對齊的 HTK 標籤 (問題3、4) **(進行至此)**
 - 從 HTK 標籤生成
 - 從對齊的 HTK 特徵中提取音素持續時間
 - 提取聲學特徵
 - **訓練持續時間與聲學模型**
 - **合成**

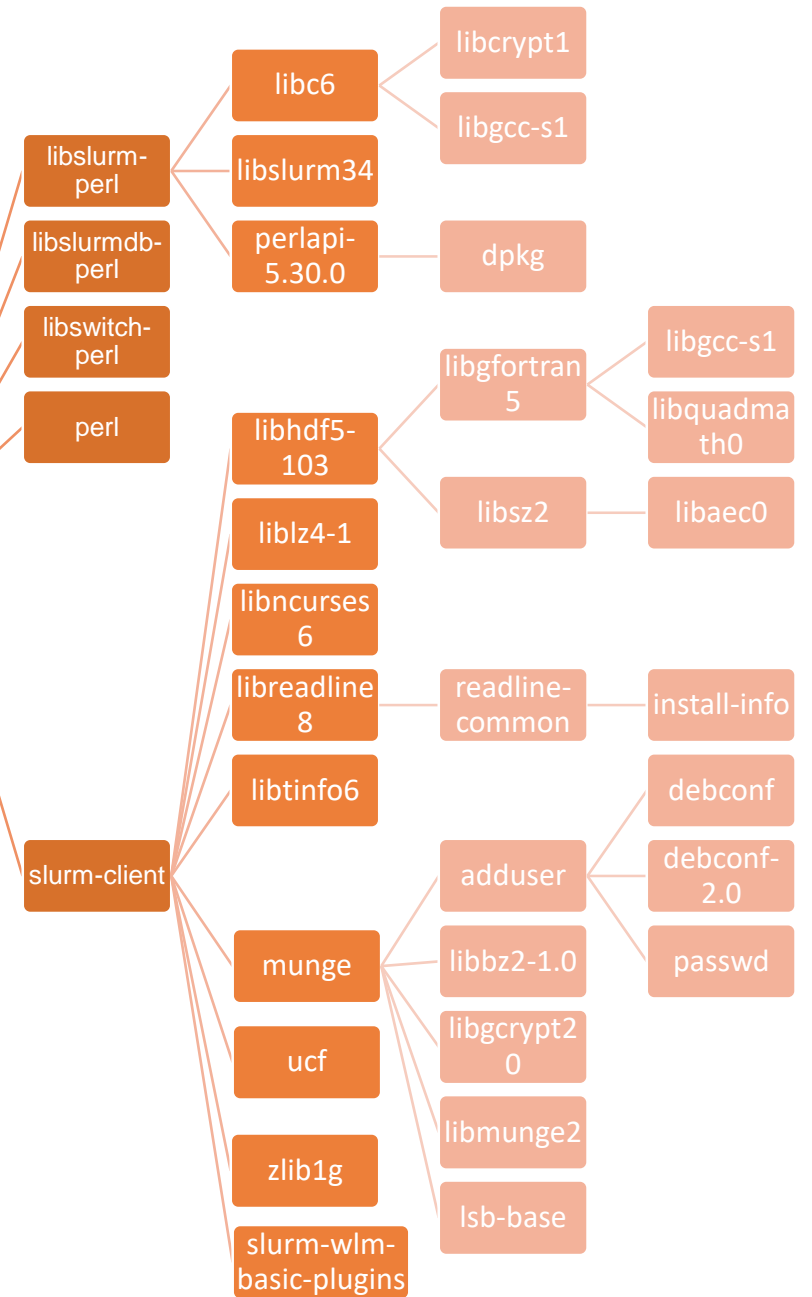
HTK：The Hidden Markov Model Toolkit，用於建構或是操作隱式馬可夫模型的工具包。應用於語音識別、語音合成、字符辨識等。

Problem1: Executing source cmd.sh error



Requires :

slurm-wlm-torque

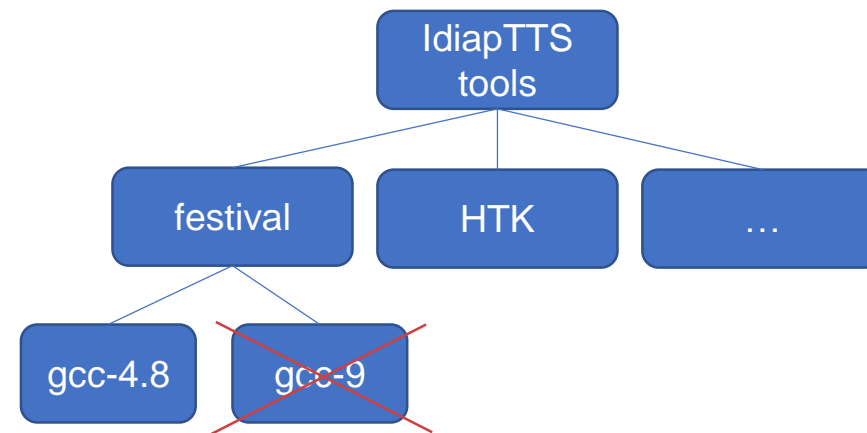


■ Problem2: Could not create library

内容待補

Problem3: Can't use festival tool

- 生成標籤時，使用到 **festival** 這個工具而報錯。
- 在 Ubuntu 20.04 版本中，官方並沒有支援 **gcc-4.8**，因此在安裝的時候我就直接選擇使用 **gcc** 最新版本進行編譯，編譯完成後發現 **festival** bin 檔損毀。
- 但因為官方沒有提供相關套件因此不好安裝，安裝完後系統仍使用預設 **gcc-9** 版本進行編譯，需使用 **alternative** 套件更改版本優先權，變更完後再次進行 **festival** 編譯。



```
(IdiapTTS) ecl@ECL-1700:~/chullin/SPSS_test/gcc-build/xgcc-4.8.5$ sudo update-alternatives --query gcc
Name: gcc
Link: /usr/bin/gcc
Status: auto
Best: /usr/bin/gcc-4.8
Value: /usr/bin/gcc-4.8

Alternative: /usr/bin/gcc-4.8
Priority: 40
```

```
(IdiapTTS) ecl@ECL-1700:~/chullin/SPSS_test/gcc-build/xgcc-4.8.5$ gcc -v
Using built-in specs.
COLLECT_GCC=gcc
COLLECT_LTO_WRAPPER=/usr/lib/gcc/x86_64-linux-gnu/4.8/lto-wrapper
Target: x86_64-linux-gnu
Configured with: ../src/configure -v --with-pkgversion='Ubuntu 4.8.5-4ubuntu8' --with-bugurl=file:///usr/share/doc/gcc-4.8/README.Bugs
--enable-languages=c,c++,go,d,fortran,objc,obj-c++ --prefix=/usr --program-suffix=-4.8 --enable-shared --enable-linker-build-id --libex
ecdir=/usr/lib --without-included-gettext --enable-threads=posix --with-gxx-include-dir=/usr/include/c++/4.8 --libdir=/usr/lib --enable
-nls --with-sysroot=/ --enable-clocale=gnu --enable-libstdcxx-debug --enable-libstdcxx-time=yes --enable-gnu-unique-object --disable-li
bmudflap --enable-plugin --with-system-zlib --enable-objc-gc --enable-multiarch --disable-werror --with-arch-32=i686 --with-abi=m64 --w
ith-multilib-list=m32,m64,mx32 --with-tune=generic --enable-checking=release --build=x86_64-linux-gnu --host=x86_64-linux-gnu --target=
x86_64-linux-gnu
Thread model: posix
gcc version 4.8.5 (Ubuntu 4.8.5-4ubuntu8)
```

■ Problem4: Failed to generate label

内容待補

■ (附錄) Statistical Parametric Synthesis

- 統計參數語音合成(SPSS) 主要比較對象為 unit Selection Synthesis。
- 生成語音所需要的聲學參數，然後透過數學方法恢復語音，其中包含了文本分析、參數預測 (聲學模型)、聲碼器分析/合成 (聲碼器)三部分。
- SPSS 系統可以看作是 ASR 的鏡像系統：ASR 系統嘗試使用機器學習模型將語音從聲學特徵轉換為一串單詞，而 SPSS 系統嘗試使用機器學習模型將一串單詞轉換為聲學特徵或直接轉換為聲波波形。
- ASR 和 SPSS 系統通常都使用大量語音數據及其轉錄進行訓練，從而產生一組描述語音數據統計特徵的參數，因此稱為“統計參數”語音合成。
- 首先從語音數據庫中提取語音的參數表示，包括頻譜和激勵參數(mfcc, lsf, f0..等)，然後使用一組生成模型 (例如，HMM) 對其進行建模。最大似然 (ML) 標準通常用於估計模型參數，最後從語音的參數表示中重建語音波形。

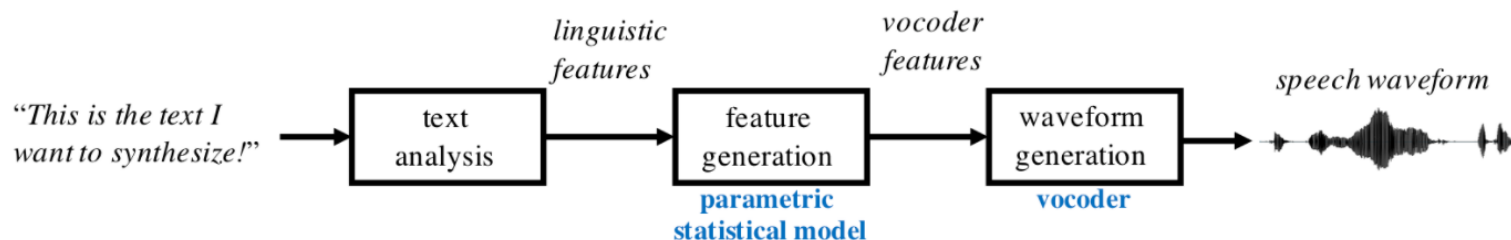


Fig. 9: A schematic view of an SPSS system

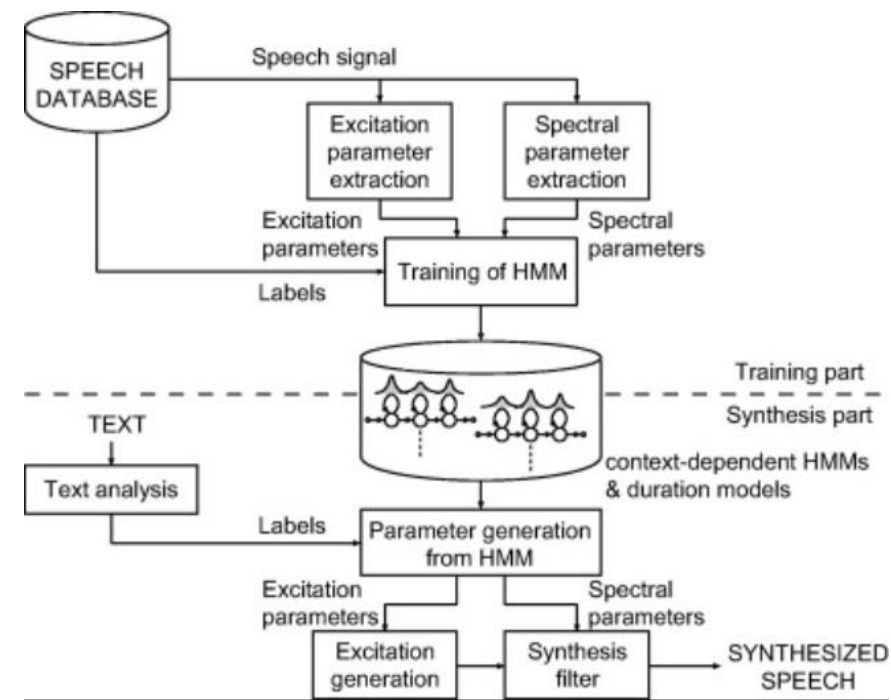


Fig. 8: Block-diagram of HMM-based speech synthesis system (HTS) [3]