

# 真假音之聲學特徵分析與轉換

## Feature Analysis and Conversion between Chest voice and Falsetto

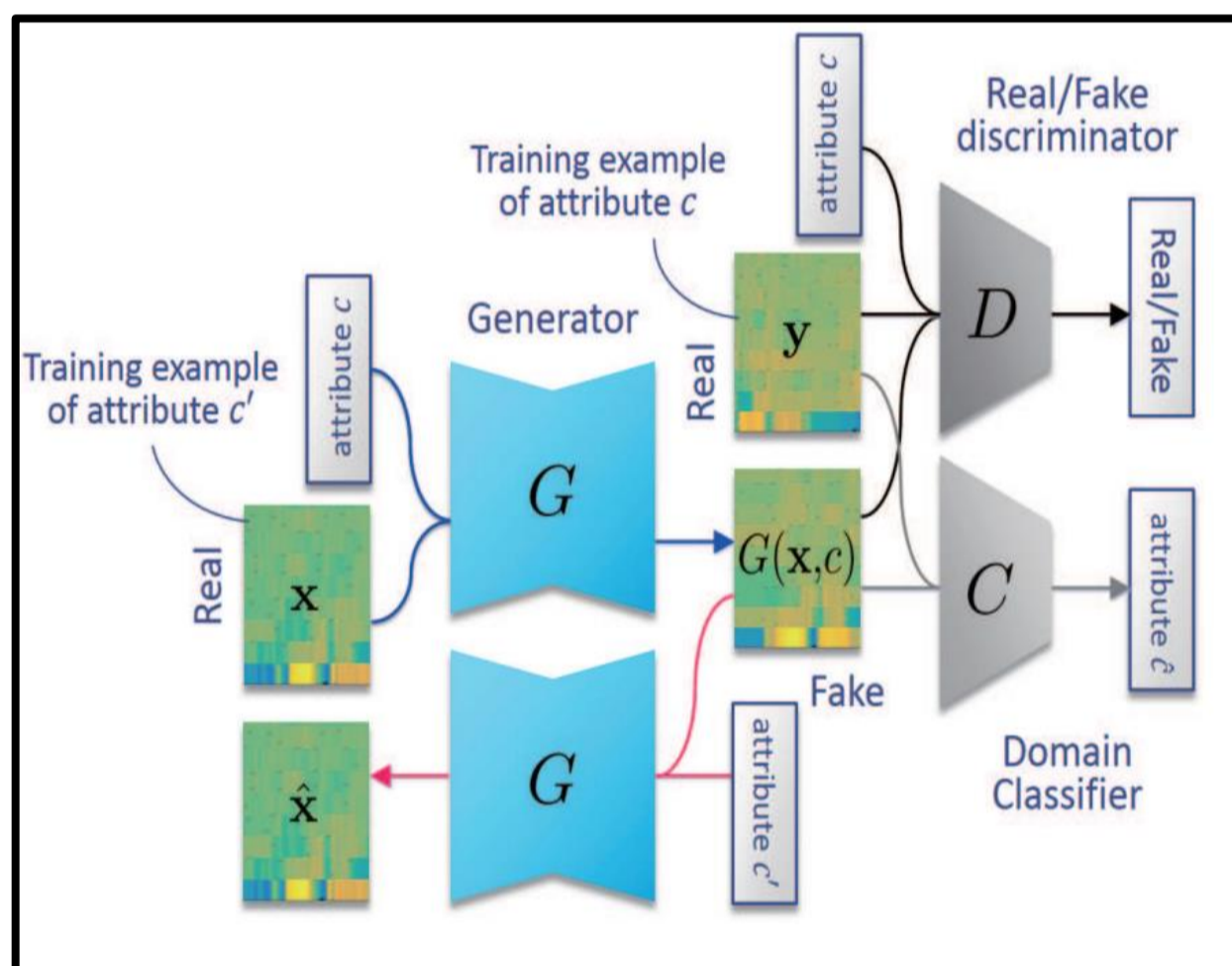
組別：A183 組員：蘇東誠 鄭璟玢 黃凱琳 指導教授：劉奕汶

### 摘要

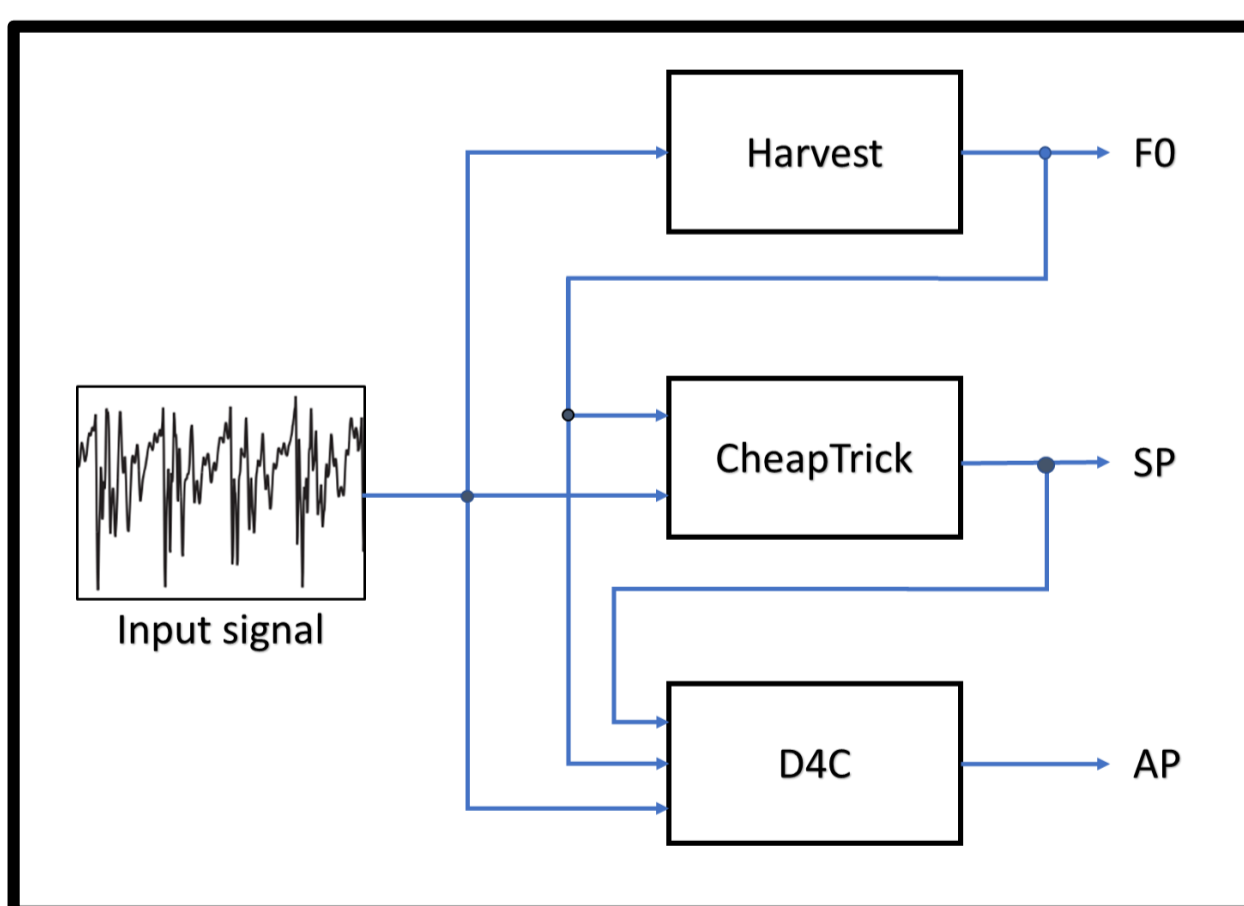
此研究主要著重在流行樂中常見的音色—真音與假音之間的轉換。我們此次的研究是聲碼器 WORLD VOCODER[2] 與語音轉換模型 StarGAN-VC[1] 的融合應用。我們利用此 Vocoder 抽取音檔的重要參數： $f_0$ 、SP、AP，將真音與假音的 SP 與 AP 作為 StarGAN 的輸入，進行轉換後再由 Vocoder 重新合成為目標音檔。資料庫的部分，我們也自行錄製了演唱流行歌曲的真音與假音共約 30 分鐘。在得到最終轉換結果後，我們對大眾進行了聽感測驗，幫助我們以客觀角度分析此次的研究成果。

### 系統架構

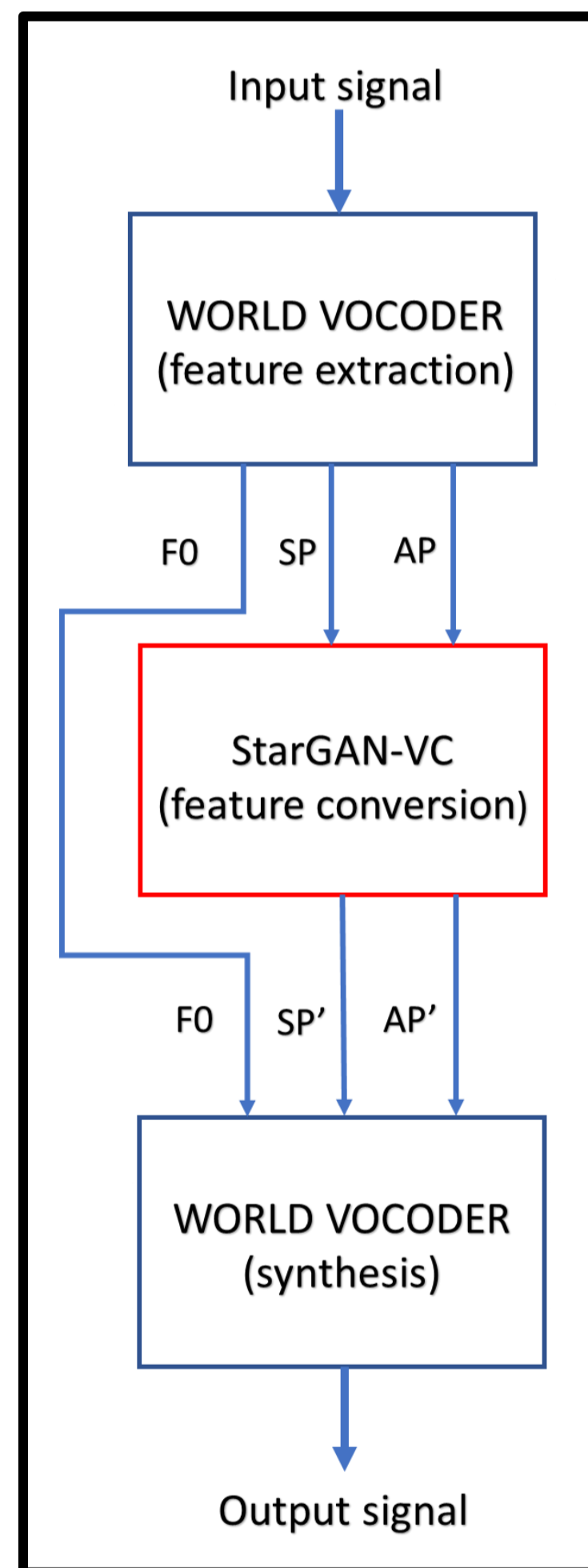
我們的真假音轉換系統為一生成對抗神經網路與一聲學模型組合而成，詳細架構見下方架構圖。



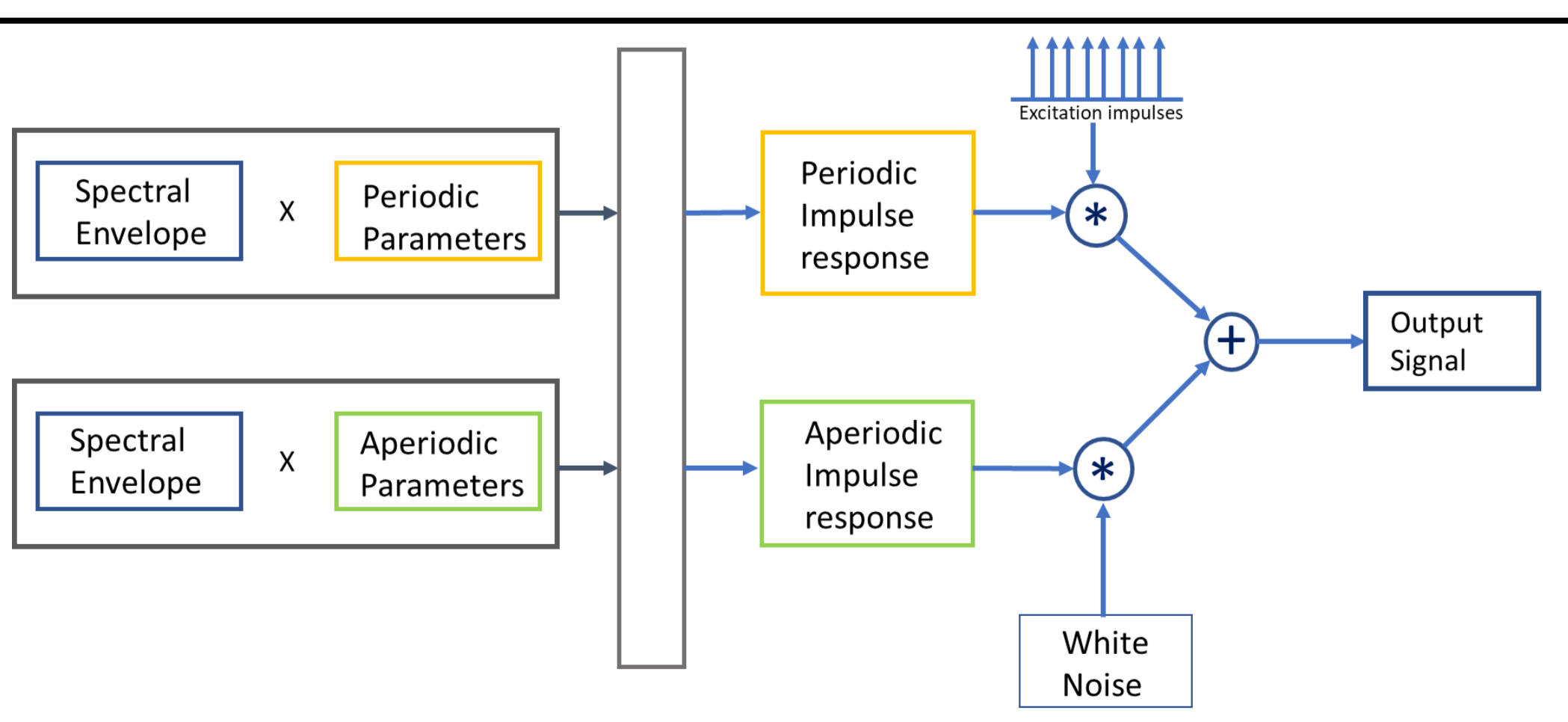
StarGAN-VC[1] 神經網路模型



World Vocoder 聲學模型: 分析方法



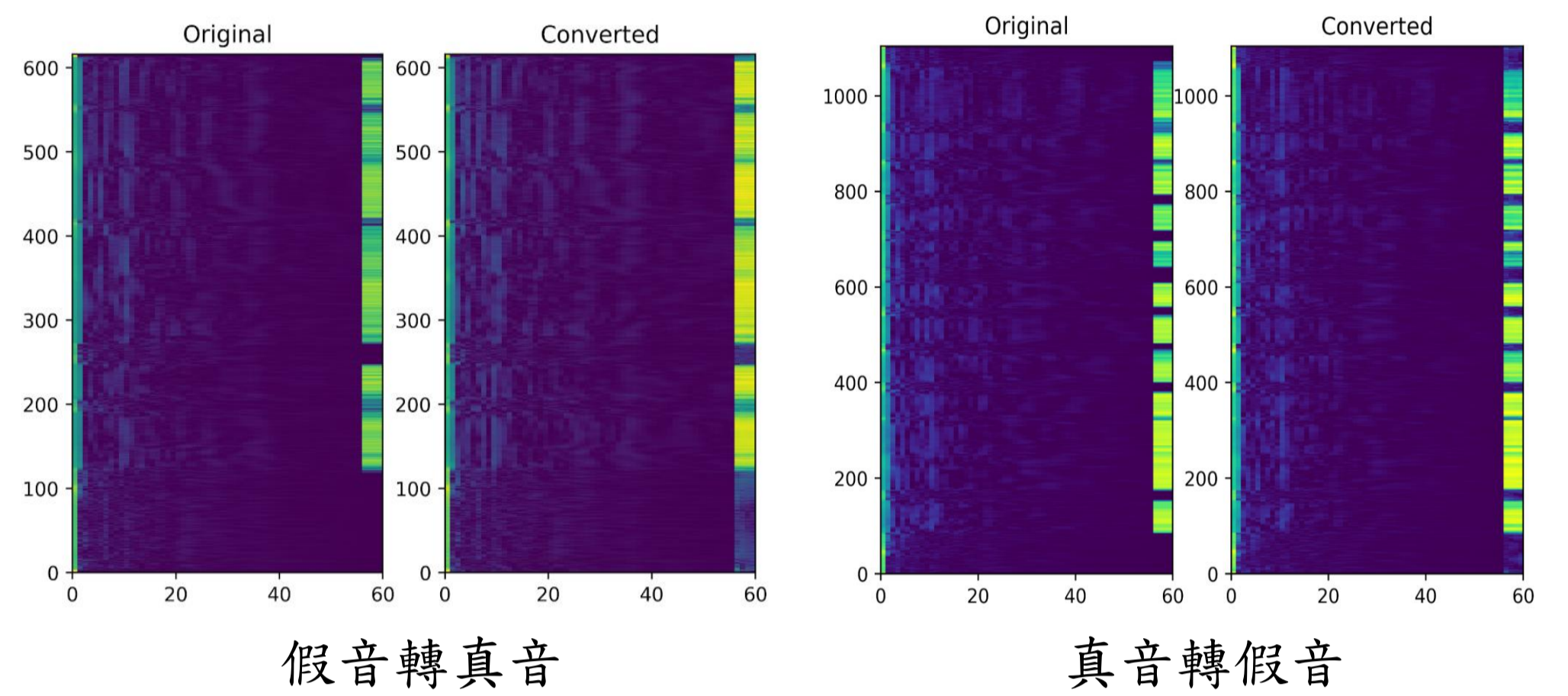
真假音轉換模型



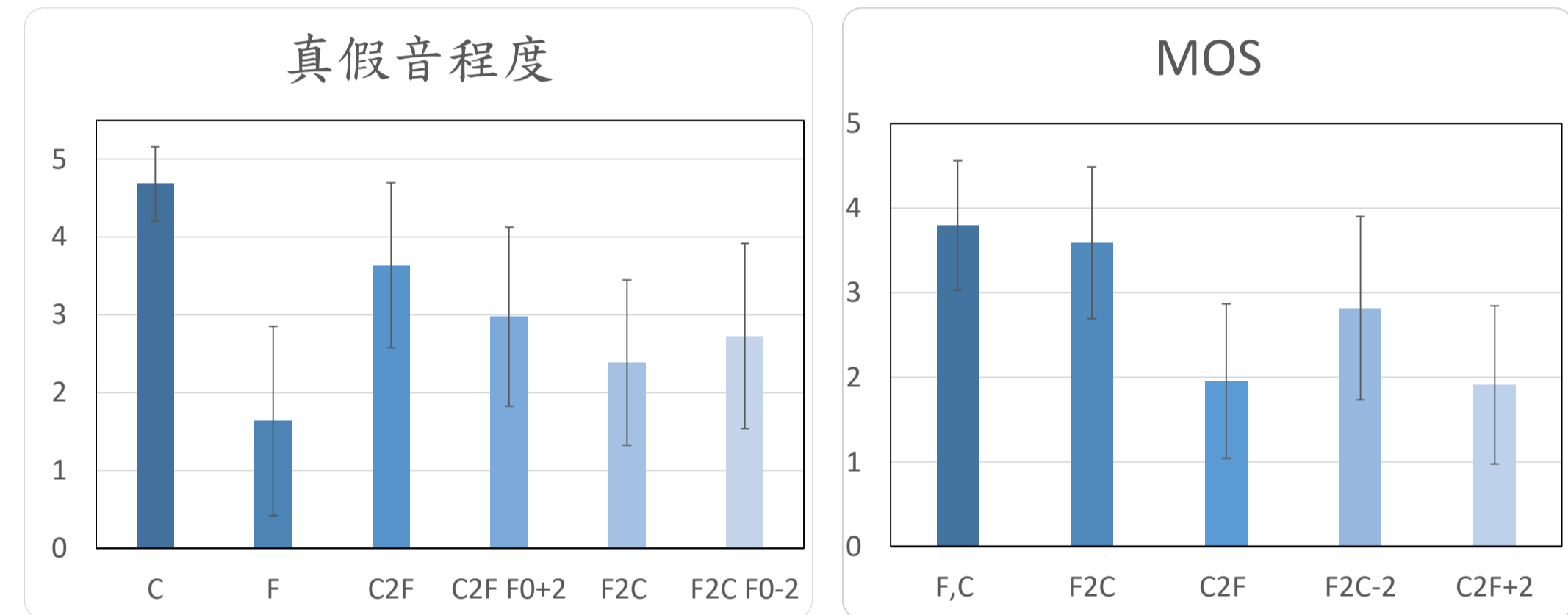
World Vocoder 聲學模型: 合成方法

### 結果

#### a. 聲學特徵轉換結果



#### b. 轉換效果與音質評估



### 結論

由上面的結果可以發現，我們轉換音色的效果是可以被人辨別的，或許音色仍無法達到原該音色音檔的水準，但音色的部分轉假轉真已經可以讓人產生混淆的效果。此外可從結果部分得知，利用此 vocoder 調整音高確實會改變到聲音的音色聽感，調高會讓音色變假，調低則會讓聲音變真。音質的部分，假轉真的部分音質較佳，真轉假的音質則相對較差。

基於音高  $F_0$  會對音色造成影響，則其的關聯性將會是未來要探討的目標，由這次的專題實驗結果我們推論在不同  $F_0$  下需要生成不同類型的 SP, AP 特徵才能讓真音的特色維持，讓人持續定義它是真音。而底下隱含的 SP, AP 特徵是否有與  $F_0$  之間有一定規則？這個規則是否能用機器學習的方式預測？或者是用非機器學習的方式就能實踐？這些會是我們下個需要解決的問題。

### 參考資料

1. Kameoka, Hirokazu, et al. "Stargan-vc: Non-parallel many-to-many voice conversion using star generative adversarial networks." 2018 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2018.
2. M. Morise, F. Yokomori, and K. Ozawa: WORLD: a vocoder-based high-quality speech synthesis system for real-time applications, IEICE transactions on information and systems, vol. E99-D, no. 7, pp. 1877-1884, 2016.