

# Status Report

---

Student: 周長毅、方泰翔、陳憲億

Advisors: Chingwei Yeh and Tay-Jyi Lin

# Outline

- BT mic for DVC SoC & vMic on DVC
  - 目前進度：
    - T90 DVC 環境建置
    - 燒錄喉麥程式以及參數(進行中)
      - 將 wav 檔取出並轉換成文字檔 (.txt)
- 準備Demo事宜(泰翔)
  - 製作海報
  - 製作影片
- 未來規劃
  - 一、使用vMic當做 input 並利用 DVC 演算法在 T90 DVC 上轉換
    - 1. 直接使用vMic輸入T90並從喇叭播放轉換音檔
    - 2. 實現real-time轉換
  - 二、將步驟一轉換成功的聲音利用藍芽連接至手機

# Demo 影片



# 震動式麥克風

## 背景介紹

聲音以空氣作為媒介傳遞，當聲音傳遞至空氣傳導麥克風之腔體振膜，經線圈或隔板利用電磁感應原理產生電流或電壓訊號即可用電子系統進行後續如傳輸、儲存等處理。因此以空氣作為傳遞媒介的麥克風若無經過特別處理，則不相關的音波將與目標音訊同時被收錄，降低收音的品質。(比如通話時車流及路人交談的聲音、或騎車時風切聲等)

而常見收音降噪的方法包括以下兩類：

- 「指向型麥克風」，利用機械設計收錄特殊角度的音波 (如圖二中之心型指向、全指向、槍型指向等麥克風設計)
- 「麥克風陣列」與「beamforming」等高階訊號處理技術 (如圖三) 的抗噪麥克風，由於數位技術突飛猛進及麥克風製造成本大幅降低，在市場上如雨後春筍般出現 (如 Apple AirPods Pro)

此兩種方式主要原理皆是針對發音源收音，同時阻斷或抑制其他源訊的聲音達到抗噪的目的，雖然可以有不錯的抑制雜訊效果，但多數環境下仍無法完全去除由四面八方反射、散射進入的雜訊干擾。

圖二 指向型麥克風

圖三 陣列麥克風 beamforming 技術

## 震動式麥克風雛形設計

此設計將加速度計作為麥克風，收錄乾淨且不受周遭環境噪音干擾的語者語音震動訊號，其相較於傳統麥克風有更強的抗噪效果，並藉由基於深度神經網路模型之語音轉換系統改善收錄之原始喉部震動訊號，使震動訊號缺失的高頻成分以及無法收錄之唇齒震動能夠透過神經網路運算進行補償。

此雛形設計包含以下兩關鍵步驟：

**步驟一：**利用裝設在說話者頸部之加速度感測器 (accelerometer; 352C33) 擷取說話時喉部之振動訊號，經由資料擷取設備 (圖四中 NI cRIO 9047 & NI-9231 DAQ) 傳入此雛形設計的 AI 擬真計算器 (圖四左的筆記型電腦)

**步驟二：**以筆記型電腦執行 AI 擬真計算，將擷取振動訊號即時轉換為語音訊號 (圖五(a))

其中用以仿真的 AI 技術可以是一個標準的神經網路，其訓練方法如圖五(b)：先錄製同步之「直接感測源訊」及其在特定之聲場環境下 (如圖五(c)之隔音室環境) 經空氣傳輸介質共振、反射、散射等通場效應產生之「目標音訊」；此配對之同步「直接感測源訊」與「目標音訊」即可作為神經網路之訓練資料。



圖四 震動式麥克風系統測試環境



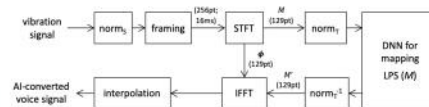
圖五 (a) 利用直接感測源訊感測及 AI 轉換之收音裝置，(b) AI 神經網路訓練方法，(c) 於隔音室錄製同步震動訊號與語音之現場照片

## AI 擬真運算

首先將振動訊號正規化 (normalization;  $norm_1$ )，避免訊號能量過大或過小影響轉換品質；接著將振動訊號切成連續的訊號框 (framing; 本實施例採用 50% 重疊、長度為 16ms 之訊號框) 進行後續處理：

- 每一振動訊號框皆經過「短時傅立葉轉換 (short-time Fourier transform; STFT)」產生對應之 magnitude 及 phase，其中 magnitude 經「對數轉換」產生「對數功率頻譜 (log power spectrum; LPS)」。
- 該振動訊號之 LPS 以訓練資料為基礎進行正規化 ( $norm_2$ ) 後，經由一深層網路 (deep neural network; DNN) 映射至語音訊號之 LPS。
- 該映射之 LPS 經反正轉化 ( $norm_3$ ) 及指數轉換，並搭配原振動訊號之 phase，以「反傅立葉轉換 (inverse Fourier transform)」產生映射之音訊訊號。

最後，映射完且重疊的連續音框以內插 (interpolation) 方式產生最後輸出之語音訊號。



圖六 轉換震動訊至語音之 AI 擬真系統方塊圖

## 轉換結果

下圖為補償後震動麥克風訊號 (CVM) 與空氣傳導式麥克風錄製語音 (ACM) 結果比較，可見不論 inside/outside 之補償後震動麥克風訊號皆可將震動訊號 (VM) 中組成子音主要的高頻噪音部分缺失進行補償，以得到無干擾且不失真的麥克風訊號。

