

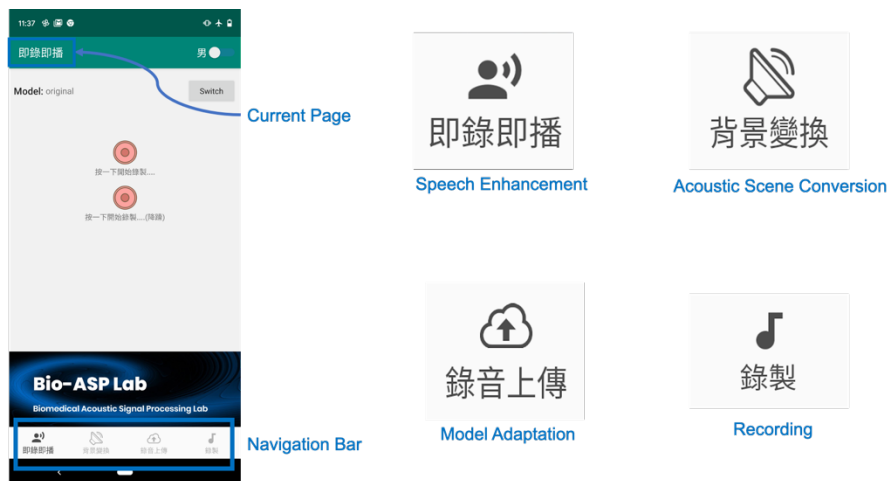
● 簡介

在這項研究中，我們提出了一種基於深度學習的語音信號處理手機應用程式，稱為 CITISEN，它支持三種功能：語音增強（SE），聲音場景轉換（ASC）和模型適應（MA）。就 SE 功能而言，CITISEN 可以有效減少語音信號中的噪聲成分，從而提高其清晰度和理解度。對於 ASC 功能，CITISEN 會將當前背景聲音轉換為聽起來像不同背景的聲音。最後，當遇到未知的說話者或未見過的噪聲類型時，CITISEN 中的 MA 功能可以有效地調整具有很少音頻文件的 SE 模型。調整過的 SE 模型可用於對嘈雜言語內容進行增強。實驗結果證實了這三種功能在客觀評估和主觀聽力測試方面的有效性。令人鼓舞的結果表明，已開發的 CITISEN 手機應用程式可以潛在地用作各種與語音相關的服務的前端處理器，例如語音通信，助聽設備和虛擬實境耳機。

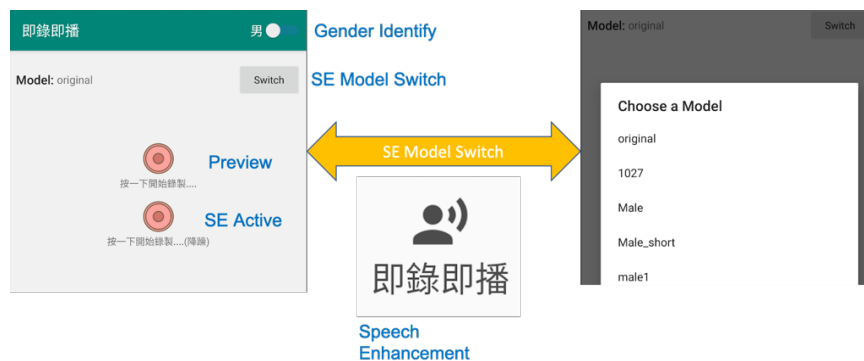
● 用戶界面和操作方式

CITISEN 中的四個主要頁面

CITISEN 應用程式有四個頁面：“即錄即播”，“背景轉換”，“錄音上傳”和“錄製”。該應用程式的左上角和底部分別列出了每個頁面的名稱和導航按鈕。



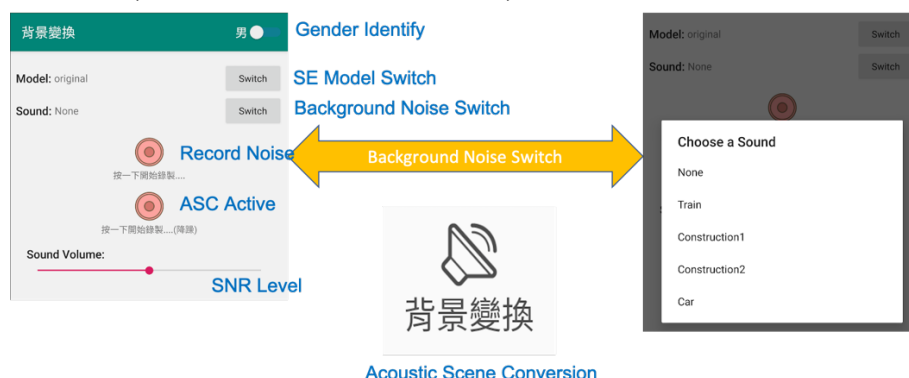
◆ 即錄即播頁面 (Speech Enhancement)



對於“即錄即播”頁面，用戶首先指定他/她的性別身份。然後，通過按下“SE 模型切換”按鈕，用戶可以從既存的模型列表中選擇一種合適的 SE 模型。CITISEN 提供了一些我們使用自己收集的語音數據進行訓練的原始 SE 模型。用戶還可以運行 MA 來生成

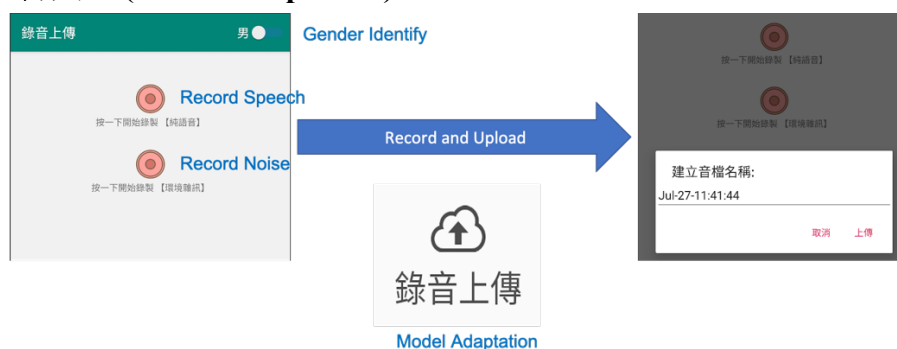
調整後的 SE 模型並將其另存為新的 SE 模型。然後，通過按 SE 按鈕，將嘈雜的語音轉換為在線乾淨的語音。

◆ 背景轉換頁面 (Acoustic Scene Conversion)



在“背景轉換”頁面中，CITISEN 在增強的語音上混合了聲音場景，以生成具有轉換後的聲音場景的新語音信號。“背景轉換”頁面上有一個“記錄噪聲”按鈕，用戶可以通過該按鈕記錄和保存 ASC 的噪聲信號。同時，該頁面具音量調整，用戶可以調整背景噪聲的音量大小，並相對應地指定轉換後語音的 SNR 級別。如要改變聲音場景，用戶首先可按“SE Model Switch”按鈕選擇 SE 模型。然後，通過按“Background Noise Switch”按鈕，如左側所示，將彈出一個聲音場景選擇窗口，並列出所有聲音場景選項，如右側所示。用戶可以選擇想要的 ASC 場景，然後語音連同轉換後的聲音場景將會生成。

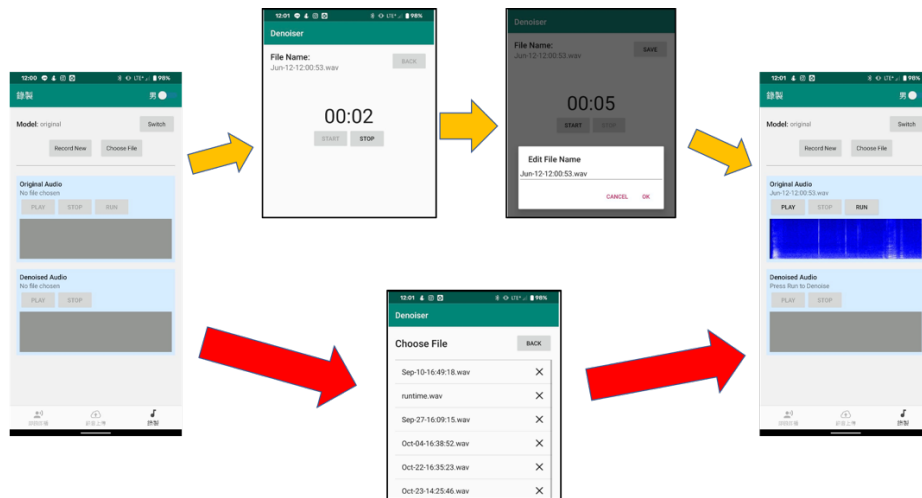
◆ 錄音上傳頁面 (Model Adaptation)



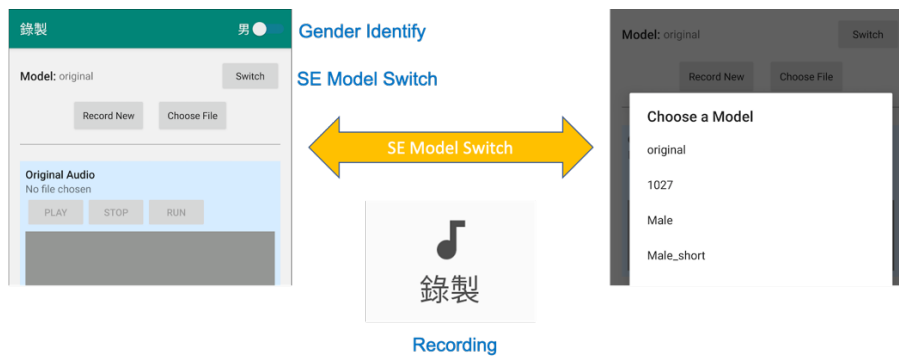
如左側所示，有兩個文件上傳按鈕：“錄製噪音”和“錄製語音”。通過按下其中之一的按鈕，用戶可以錄製純噪聲或說話者的語音信號，並將錄製的音頻上傳到我們的伺服器。如左側所示，當要開始錄製，用戶只需簡單的按下其中一個按鈕即可。如右側所示，當完成錄製後，再次按一次按鈕，CITISEN 將彈出一個提交窗口。提交窗口將要求用戶命名音頻文件，然後音頻將發送到伺服器。在接收到音頻文件後，伺服器將通過錄音的音頻數據來微調原始 SE 模型來估計調整後的 SE 模型。音頻文件的名稱還將用於命名調整後的 SE 模型，該模型隨後從伺服器發送到手機設備，並顯示在“即錄即播”和“背景轉換”的頁面中。因此，用戶可以使用調整後的 SE 模型運行 SE 和 ASC 功能。

◆ 錄製頁面 (Recording)

A. 錄製或加載保存的音頻文件



B. 選擇模型進行 SE



C. 通過頻譜圖展示經過處理的語音



“錄製”頁面供用戶錄製當前環境的語音和噪音，並保存增強或轉換後的音頻文件。在“即錄即播”和“背景轉換”頁面，用戶可以立即在線上收聽增強或轉換後的語音。另一方面，“錄製”頁面允許用戶稍後在處理過的音頻文件上可保存和回放。用戶首先錄製（圖 A 中的上方路徑）或加載現有音頻文件（圖 A 中的下方路徑），然後按“SE Model Switch”按鈕。然後，將彈出 SE 模型選擇窗口，如圖 B 的右側所示。通過選擇合適的 SE 模型，然後按下運行按鈕（如圖 B 的左側所示），將生成增強的語音。CITISEN 具有演示兩個頻譜圖的功能：噪聲頻譜圖和增強語音頻譜圖（如圖 C 右側所示），以便用戶可以直接地查看 SE 結果。除了這兩個圖外，用戶還可以在頻譜圖上按“播放”和“停止”按鈕來播放和收聽原始和處理的音頻文件。