

簡介

我們提出了一種基於深度學習的語音信號處理應用程式，CITISEN。CITISEN 支援三項功能：語音增強（Speech Enhancement, SE），聲音場景轉換（Acoustic Scene Conversion, ASC）和模型適應（Model Adaptation, MA）。就 SE 功能而言，CITISEN 能有效降低語音信號中的噪音，讓人在雜訊環境中聽懂語音，並保持語音品質。對於 ASC，CITISEN 轉換語音的背景聲，使得說話者如同身處另一個環境中與其他語者對話。最後在面對未知的使用者或環境噪音類型時，CITISEN 的 MA 功能可以即時調整 SE 模型，增加 CITISEN 系統的除噪能力。實驗結果顯示在客觀評估及主觀測試方面，CITISEN 都能有令人滿意的效果。這些結果也表明 CITISEN 應用程式能有效地做為各種與語音相關服務的前端處理器，例如語音通信，聽覺輔助設備和虛擬實境耳機。

用戶界面和操作方式

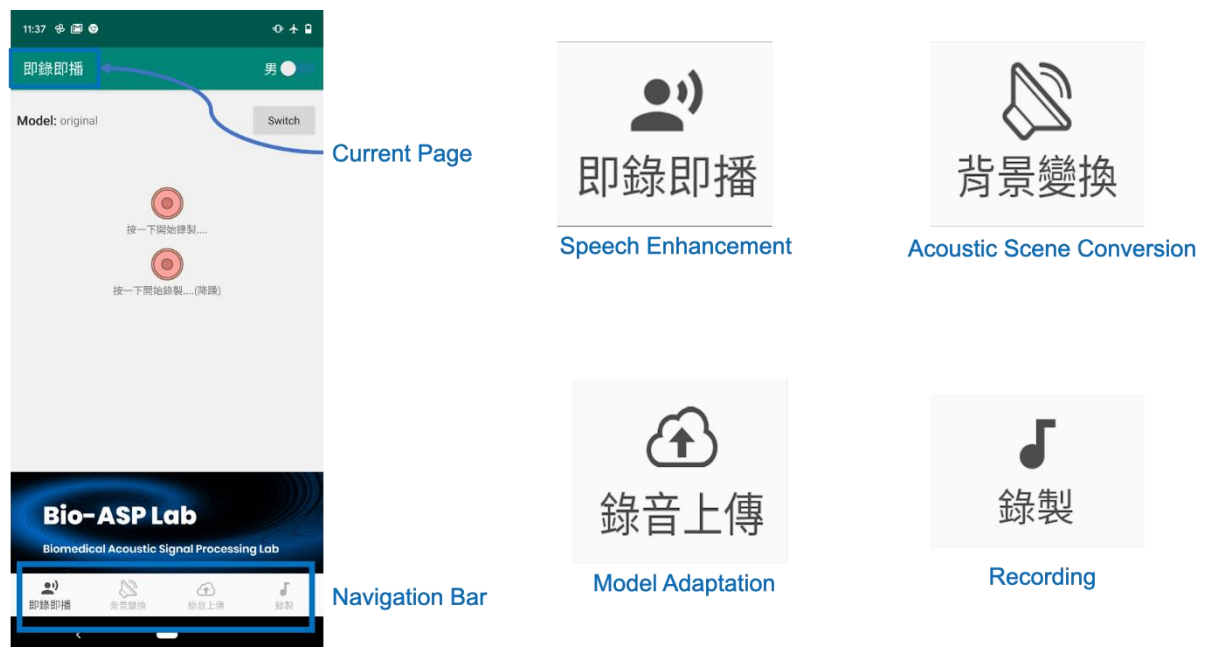


圖 1：CITISEN 應用程式有四個頁面：“即錄即播”、“背景轉換”、“錄音上傳”和“錄製”。該應用程式的左上角和底部分別列出了每個頁面的名稱和導航按鈕。

(1) 即錄即播頁面

對於“即錄即播”頁面，需首先選定用戶的性別（[圖 2](#) 右上角），再按下“Switch”按鈕，從模型列表中選擇一種合適的 SE 模型。除了現有的模型，CITISEN 也提供使用者選擇 MA 來自訂 SE 系統（在“錄音上傳”頁面），並儲存為個人化的 SE。需注意自訂系統的功能，會請使用者提供一些必要的語音資料。選擇好 SE 的模型

後，按下“按一下開始錄製...（降噪）”按鈕，CITISEN 系統將為使用者轉換嘈雜的聲音為高品質的語音。

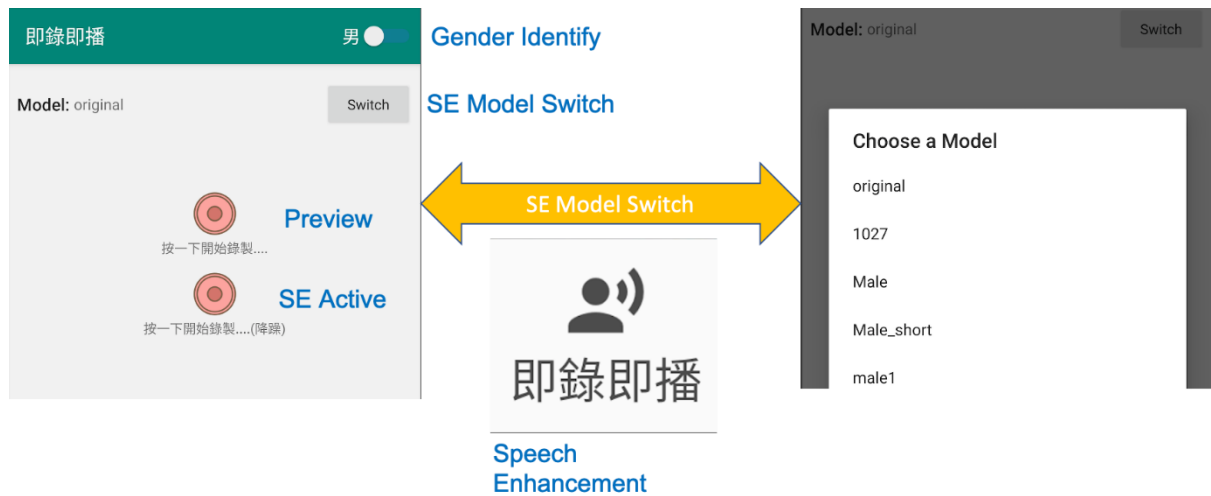


圖 2：即錄即播頁面

(2) 背景轉換頁面

在圖 3 的“背景轉換”頁面中，CITISEN 仍然能透過“SE Model Switch”選項提供使用者選擇不同的 SE 模型。而在背景變換中，使用者可以透過“按一下開始錄製”按鈕保存不同的環境聲音。如要改變聲音背景環境，可按下“Background Noise Switch”按鈕，此時將彈出一個聲音場景選擇視窗，列出所有聲音場景選項，如圖 3 右側。使用者可從中選擇背景環境；再按下“按一下開始錄製...（降噪）”以得到背景改變的語音。使用者也可在“背景轉換”頁面下方的“Sound Volume”拖曳調整背景環境音量。

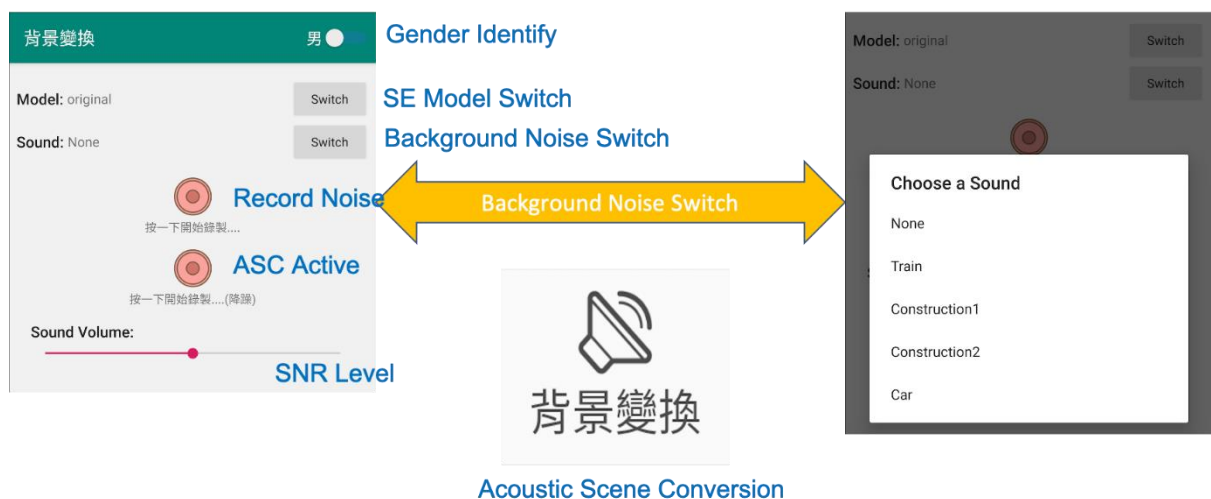


圖 3：背景轉換頁面，CITISEN 系統在除噪後的語音上混入了另一種背景聲音，讓語者有“聲歷其境”的感覺

(3) 錄音上傳頁面

“錄音上傳”頁面如圖4左側，有兩個按鈕，分別為“按一下開始錄製【純語音】”和“按一下開始錄製【環境雜訊】”。使用者可以利用這兩個按鈕，錄製純語音或環境雜訊，並將錄下的聲音上傳到遠端的伺服器中。錄音時，這兩個按鈕都是按一下開始，再按一下停止。當完成錄製後，CITISEN 將彈出一個上傳視窗，如圖4右側，提供使用者命名錄下的音檔。在遠端伺服器接收到音檔後，將透過錄下的音檔來微調SE系統。音檔檔名將做為調整後SE系統的名字，並下載到使用者的設備上，顯示在“即錄即播”和“背景轉換”的頁面中。因此，用戶可以選擇此SE系統為嘈雜聲音除噪。

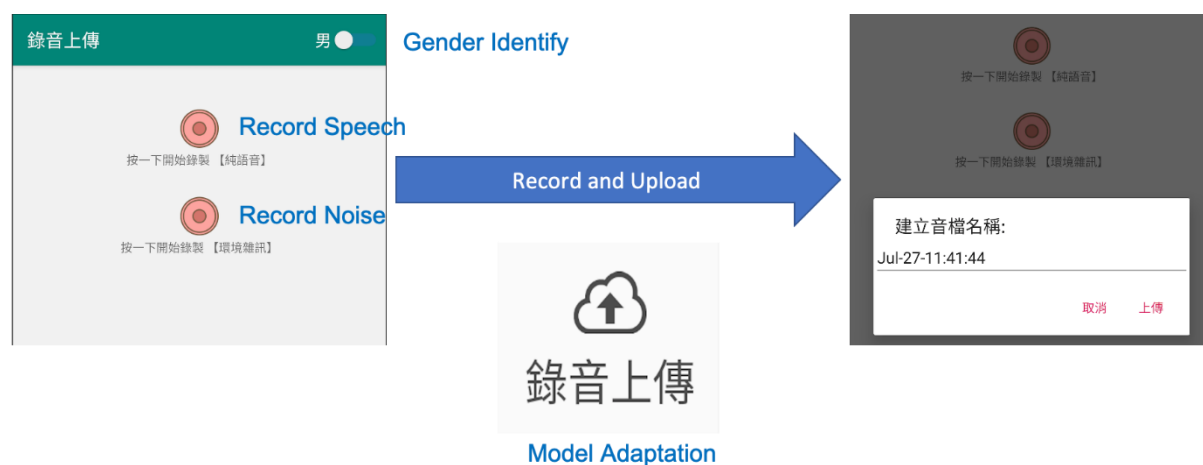


圖4：錄音上傳頁面（模型調適功能）

(4) 錄製頁面

在圖5的“錄製”頁面中，CITISEN 提供使用者儲存和播放之前錄下帶有噪音的聲音，或是除去噪音後的語音。使用方式如下：首先，使用者錄製（圖5上方路徑、圖6的“Record New”）或讀入已經儲存的音檔（圖5下方路徑、圖6的“Choose File”），再按下“Switch”按鈕，CITISEN 將彈出 SE 模型選擇視窗，如圖6右側。使用者選擇合適的 SE 系統後，按下“RUN”按鈕（圖6左側藍色區域內），產生去除噪音後的語音。CITISEN 將音檔處理前後的差異以視覺化的方式呈現（如圖7右側）。除了視覺上的感受外，使用者也可以按下“PLAY”按鈕和“STOP”按鈕來播放和停止收聽處理前（圖7右側上方）後（圖7右側下方）的音檔。

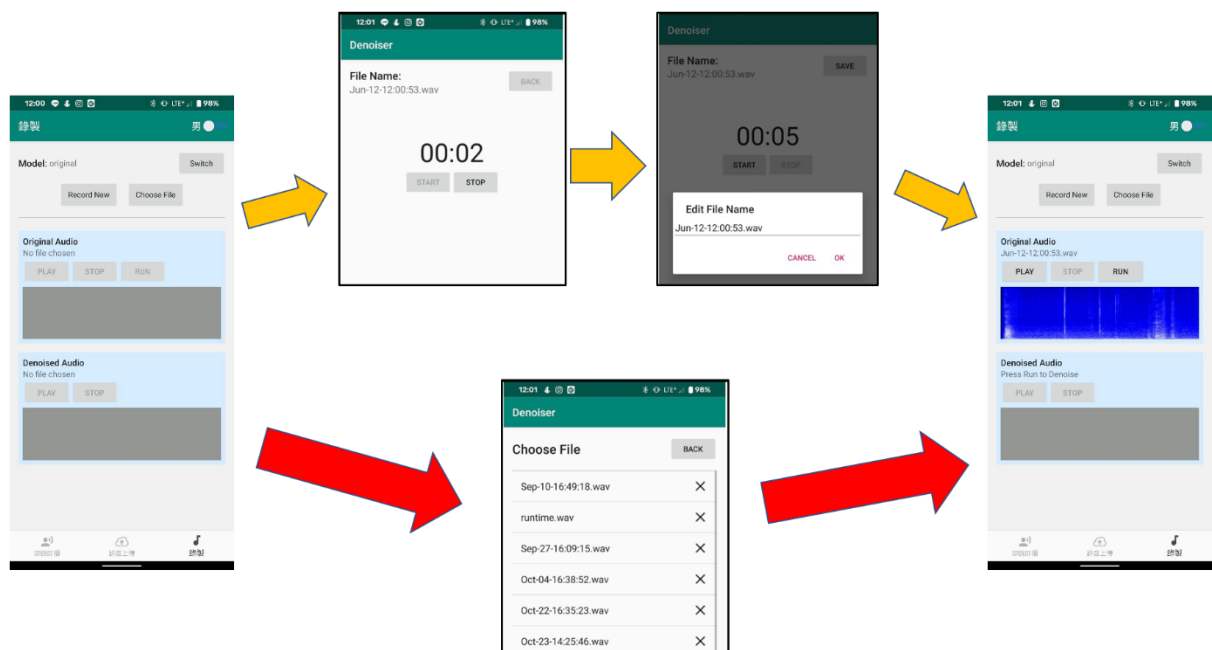


圖 5：CITISEN 提供使用者錄下帶噪音的語音，並儲存錄下的聲音及除噪後的語音。

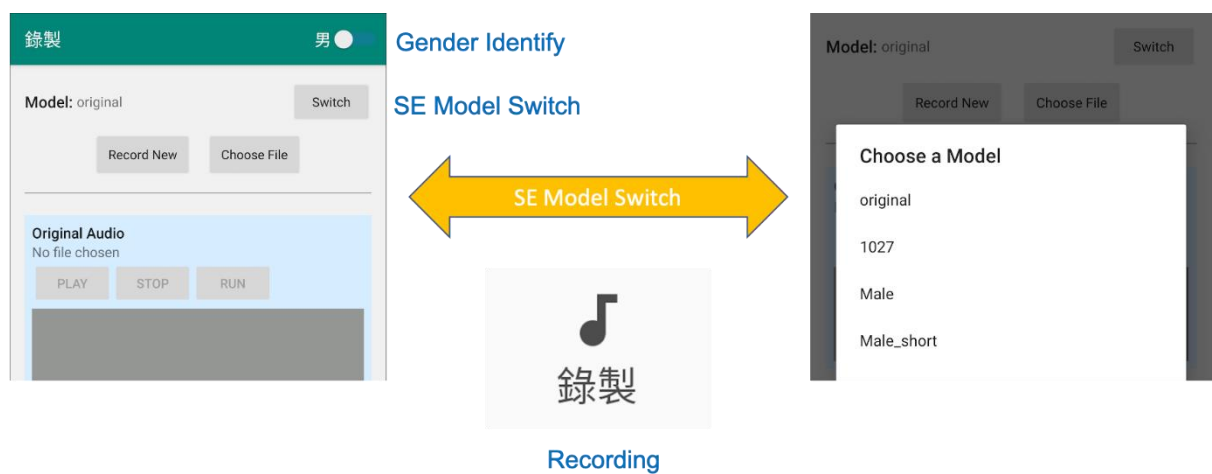


圖 6：選擇系統以去除語音中的雜訊



圖 7：視覺化顯示處理前後的語音