

死而復聲

哀悼科技研究與應用

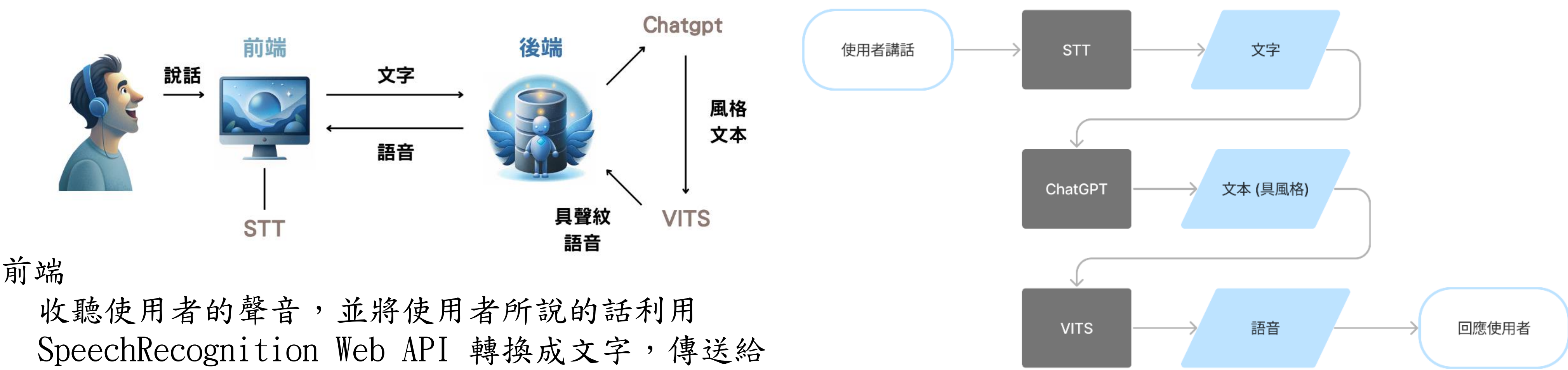
Voice Chat with Artificial Intelligence

專題組員：朱昕頤 王立宣 林柏翔 石楚城

摘要

從古至今，若是想念已故的人們，我們只能利用觀落陰等靈媒或者做夢來和已故的親人或好友交流，但隨著AI技術日新月異，為了解決人們的思親之愁，我們結合了聲音克隆模型、文本生成以及語音辨識這三個技術，開發了一個AI語音聊天程式。主要作用是與一個由我們開發的虛擬人物進行語音聊天，其中虛擬人物的聲音和說話方式可以是世界上任何人。從而達到模仿已故親人或好友，讓還在世的人能透過此網站體驗親朋好友還在世的時光。

系統流程



前端
收聽使用者的聲音，並將使用者所說的話利用 SpeechRecognition Web API 轉換成文字，傳送給後端。接收後端回傳的句子和語音，語音會和句子搭配，播放語音的同時將對應的句子打出來。

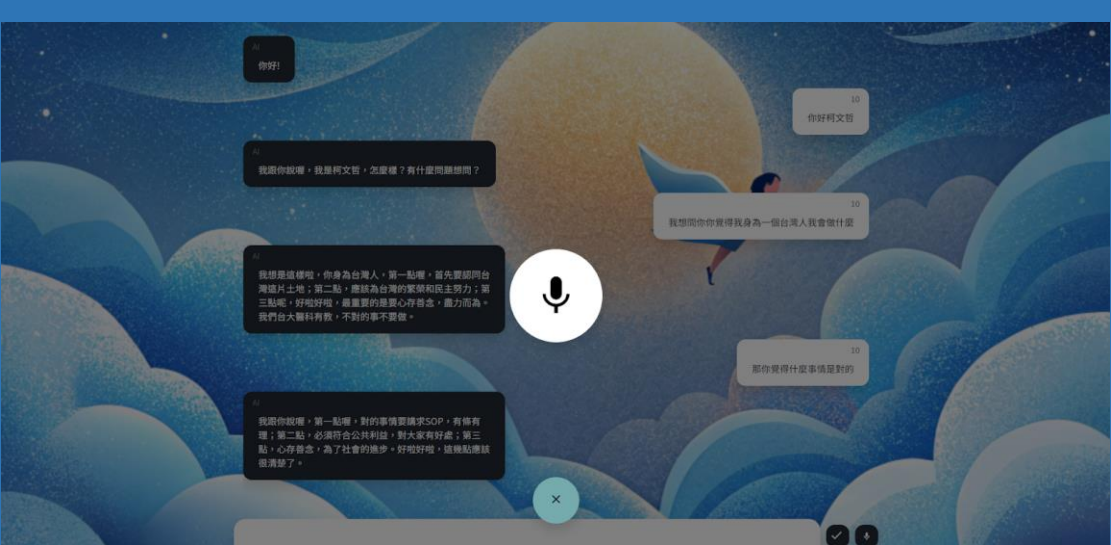
後端
接收前端所傳過來的文字，並利用OpenAI 所提供的 API 將它和準備好的「系統提示詞」合併輸入給 Chatgpt。Chatgpt 生成文字回應，將文字回應依照標點符號分割，每段輸入到已經訓練好並啟動的 VITS 微調預訓練模型做推理生成出帶有「聲紋特徵」的語音，將句子和語音一併傳送給前端。

上圖為最終的流程圖，首先我們使用者先說一段話，STT語音辨識會先幫我們把語音轉成文字，接下來將文字放入訓練好的ChatGPT進行文本生成，把生成出來的文本丟入VITS，即會得到一段語音，此段語音就會是使用者得到的回覆。

主頁面



對話頁面



接收語音

結語與展望

我們這次的專題使用了許多工具，像是聲音克隆模型、文本生成技術、語音辨識技術。從而製作得到一個AI語音聊天的應用程式。然而我們最一開始目標是想要做視訊聊天程式而非語音聊天，但是由於時間的不足再加上大部分的人臉合成模型都是屬於付費使用，再加上使用上也有不小的難度，導致我們最終沒有將此相技術加進此專題之中。未來我們將繼續嘗試加入人臉合成技術，讓我們的AI聊天程式的功能更加全面。