

2022年

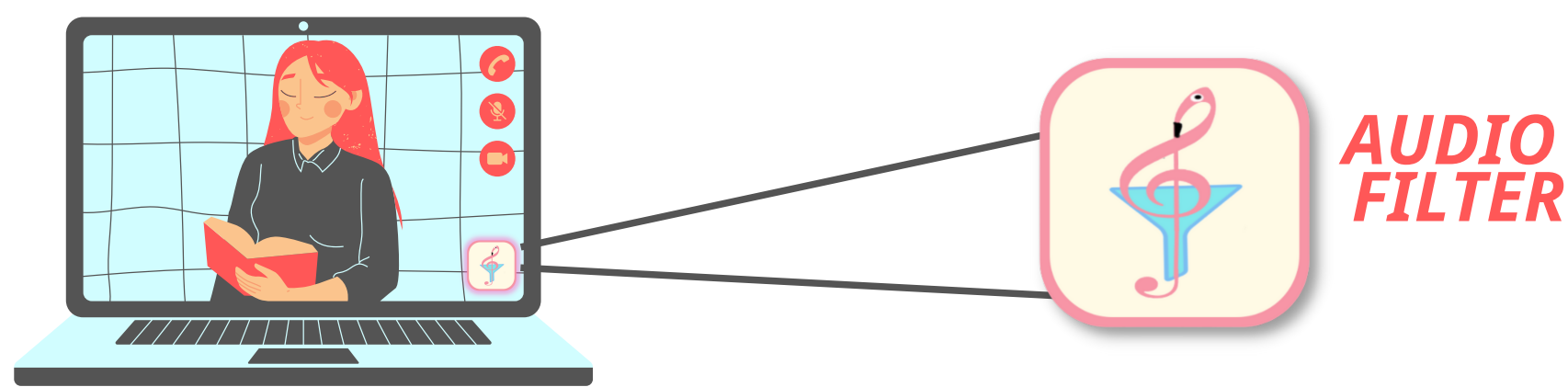
# 跨域教學聯合系列活動

## 內容：聲音濾鏡 Audio Filter

### 授課老師：張家凱 博士

## 一、摘要

後疫情時代及元宇宙的到來讓良好的線上會議體驗日趨重要，語音，更逐漸取代外貌所帶給人的第一印象。其中，冗言贅字所帶給聽者的困擾更是許多人的切身之痛。因此，我們將設計一套可應用於線上會議、編輯影片使用的輔助程式—聲音濾鏡Audio Filter，它可以刪除音訊中的冗言贅字，幫助聽者提升對於會議內容、影片內容的專注度。



於線上會議、編輯影片使用的輔助程式—Audio Filter

## 二、研究動機

後疫情時代，居家辦公、線上討論與實體會議三者兼容的混合工作模式正在崛起。根據2021年麥肯錫的研究報告指出在他們所調查的八個國家裡，2030年以前，需要這種混合工作模式的人口將比起疫情前預估超過一億人以上。此外，因應疫情遠距教學的需求也日益攀升，2021年五月，台灣進入三級警戒，多數學生也改為居家上課。

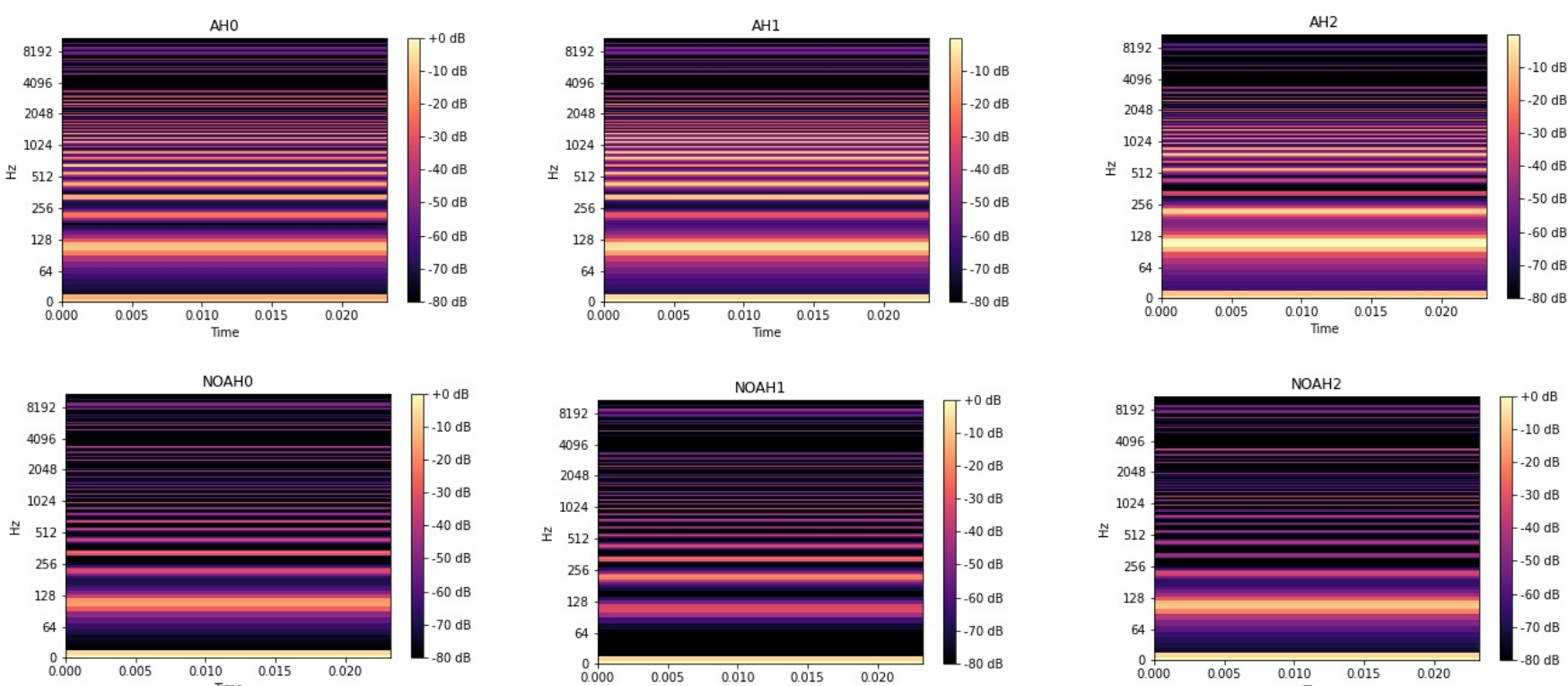
由此可見，線上會議對於人們生活的重要性變得舉足輕重，其中語音為人與人之間訊息傳遞的重要媒介，良好的音訊品質能讓聽者專注也更能有效率的傳達訊息。因此，我們希望改善目前語音傳遞時的常見的一些問題。



全球受新冠影響的停課學校狀況(2020/3)

## 三、研究方法

利用python 現有的套件librosa，可以載入聲音檔案進行STFT(短時傅立葉轉換)、MFCC(梅爾頻率倒譜系數).....等分析，分析後的資料也可以利用此套件進行視覺化的處理，使我們能更了解資料的型態以及分析後的結果，有了這些結果，我們就能訓練一個模型讓電腦幫我們進行判別以及分類。

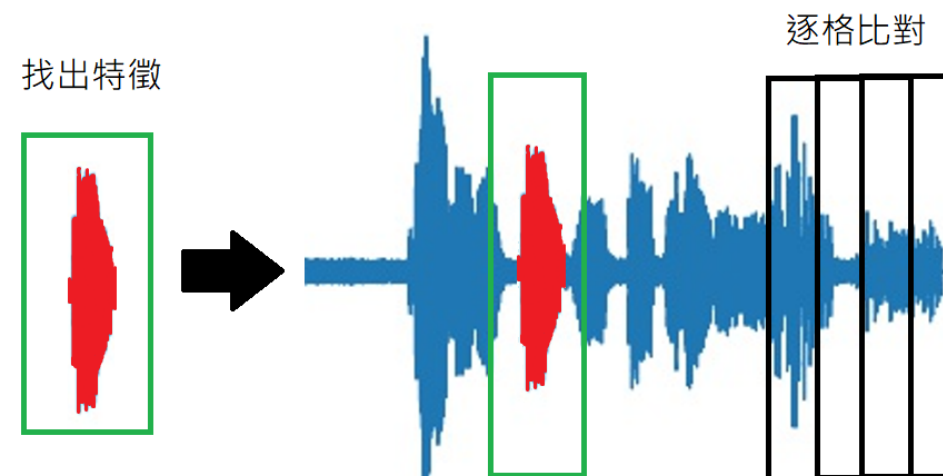


短時傅立葉轉換後的結果，可以看出第一列、第二列的不同，分別是有贅字以及沒有贅字

## 四、資料分析

訓練資料為柯文哲在公眾場合的演講音檔。使用Audacity音訊軟體，以人工方式為柯文哲使用的所有贅字貼上標籤，最終輸出格式為「時間間隔：贅字種類」文字檔。

先利用短時傅立葉轉換找出贅字的特徵，再將訓練的資料音訊切成50毫秒的長度進行逐格比對，以此方式來訓練模型



利用模型逐一比對給定的特徵音訊

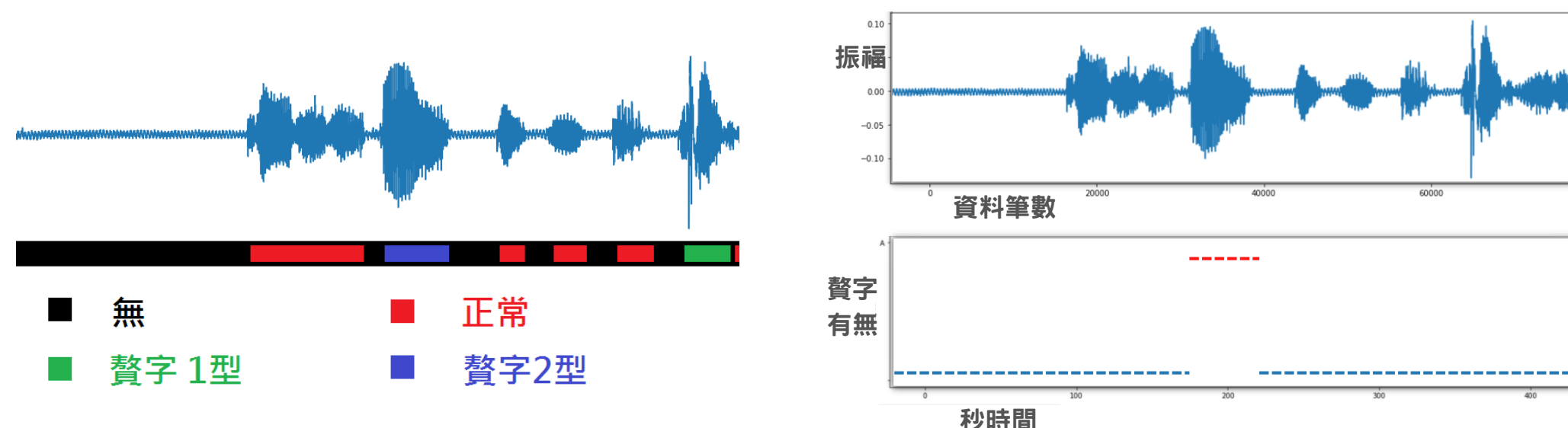
1.910027	2.094663	吼
3.217334	3.351036	吼
4.055624	4.325150	吼
4.961825	5.284408	吼
6.882463	7.050121	嘍
9.401576	9.543767	吼
12.014069	12.168993	嘍
15.012811	15.237769	嘍
15.583697	15.647364	嘍
19.562919	19.739066	吼

標籤的格式，分別是起始時間、終止時間以及贅字種類

## 五、結果與討論

將50毫秒的音訊視覺化後再進行CNN模型訓練的方法可以找出某種贅字的位置，但礙於目前訓練資料不夠多，所以可能產生擬合過度(overfitting)的問題，而且視覺化之後產生的圖片檔案非常多，需要較多儲存空間。

另一種方式則是使用Unet 架構的模型來進行語意分割(semantic segmentation)以達到找到並分類出多種贅字的目標。



以語意分割的方式分類贅字

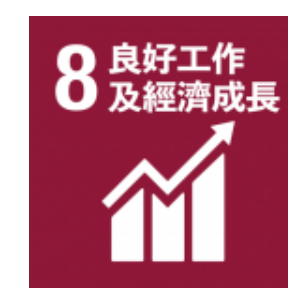
用訓練好的模型辨識非訓練的檔音檔

## 六、社會影響力

優化過的語音不拖泥帶水，句句是重點。並幫助達成多項聯合國於2015年提出的永續發展目標。



確保有教無類、公平以及高品質的教育，及提倡終身學習。過濾老師冗言贅字，提升學習專注力，提高線上教學品質。



促進包容且永續的經濟成長，讓每個人都有一份好工作。過濾線上會議發言人冗言贅字，提升遠距工作溝通效率。



建立多元夥伴關係，協力促進永續願景。增進語音傳遞品質符合第8項子目標—促進資訊傳播科技發展。

## 七、未來展望

現階段發展可供線上會議、剪片軟體使用的輔助程式，功能僅有過濾冗言贅字。隨著元宇宙時代來臨，不只影像可以透過濾鏡虛擬化，我們計畫推出適用手機通訊軟體的聲音濾鏡，除了刪除冗言贅字更能美化或改變聲音頻率(例如: 讓使人昏昏欲睡的人聲變為輕快活潑的聲調)。

現階段可以輸入一段柯文哲說話音檔，經由訓練好的模型