第五課

說

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

CUHK Jockey Club AI for the Future Project
中大賽馬會智為未來計劃

Co-organized by:

Faculty of Engineering
Faculty of Education
The Chinese University of Hong Kong

Funded by:

香港賽馬會慈善信託基金
The Hong Kong Jockey Club Charities Trust
同心同步同進 RIDING HIGH TOGETHER

內容

1

意識

口語與智能
文字轉語音(TTS)是什麼？
TTS的困難與挑戰

2

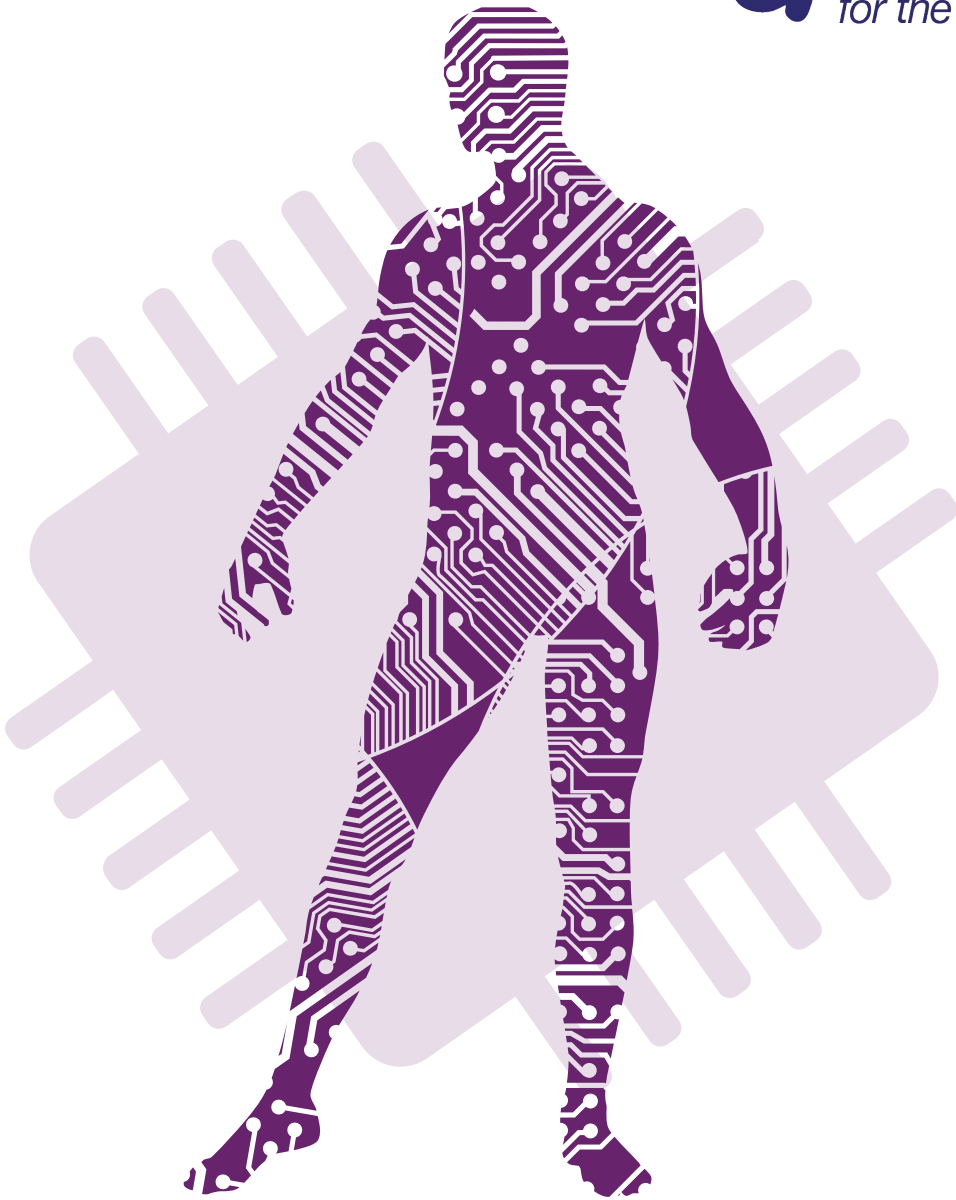
知識

文本規範化，文字到音位轉換，
韻律預測，TTS方法和評估

3

倫理

公平與安全性



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

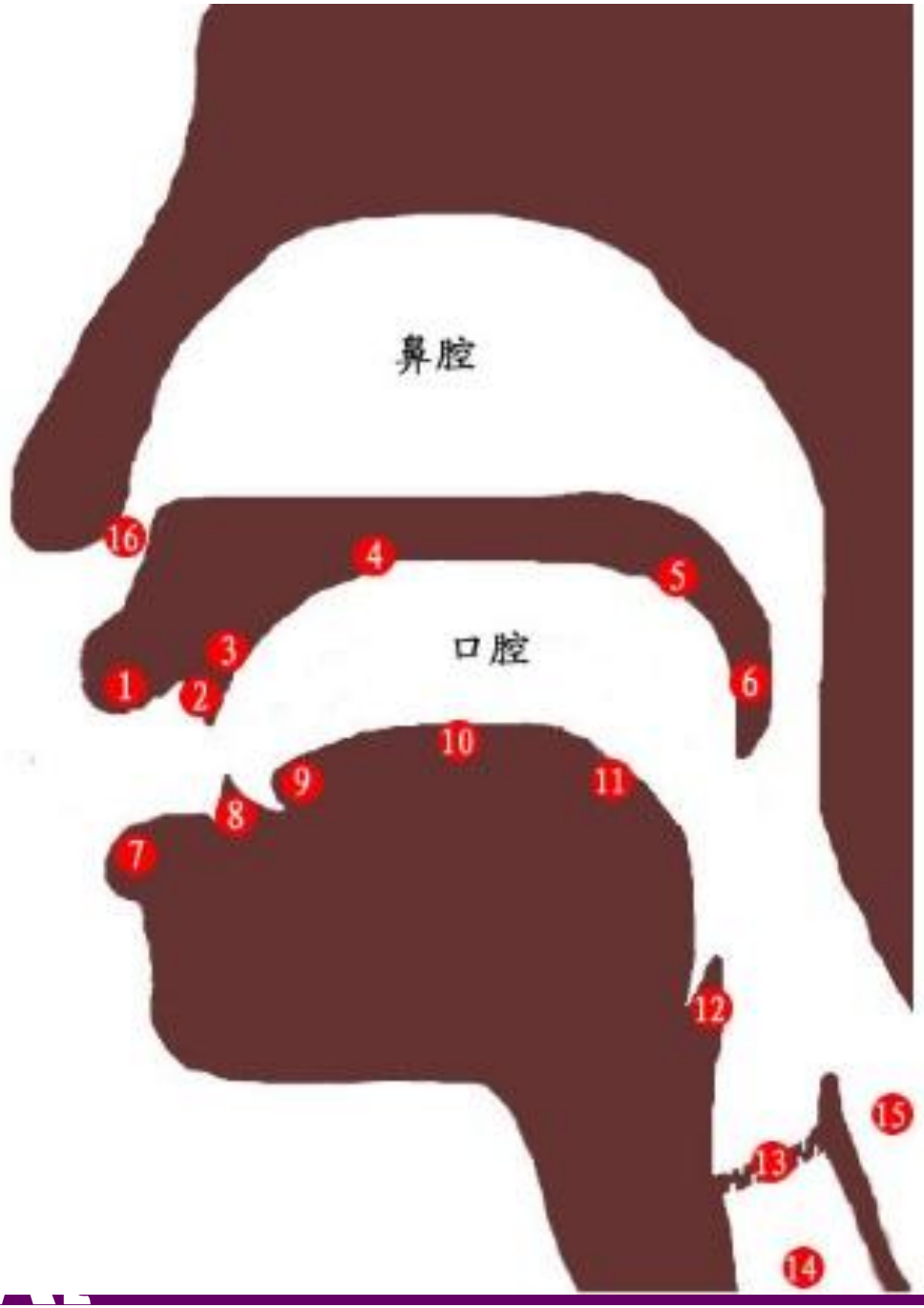
意識

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

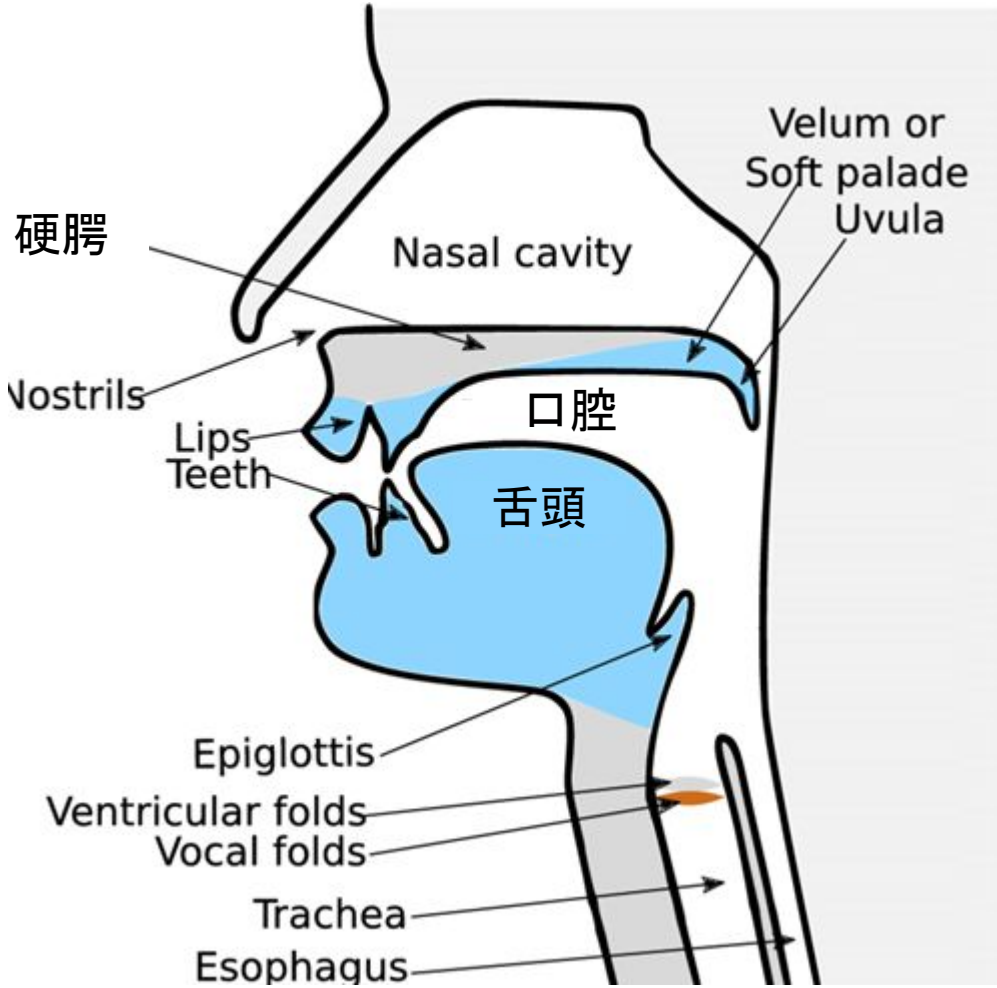
介紹

- 處理人類語言的能力是擁有智能的跡象之一
- **產生**語音的技術是AI核心技術之一
- 讓我們來看看人類是如何**發出**語音

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.



1、上唇
2、上齿
3、上齿龈
4、硬腭
5、软腭
6、小舌
7、下唇
8、下齿
9、舌尖
10、舌面
11、舌根
12、会厌（喉盖）
13、声带
14、气管
15、食道
16、鼻孔



文字轉語音(TTS)

Link: https://www.youtube.com/watch?v=_UrMV4BHY7A

文字轉語音（語音合成）取文字輸入並生成語音

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

討論

你可以列出TTS的一些應用嗎？

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

TTS的應用

例子一

幫助閱讀障礙者

資料來源:
Understood

例子二

幫助有語音障礙人士

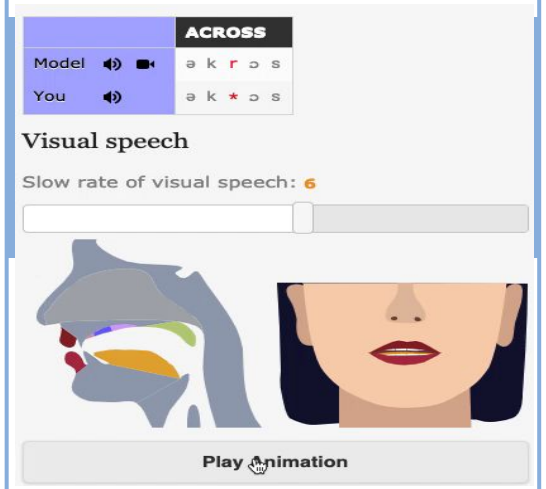
資料來源:
Bloomberg

例子三

自動公告系統

資料來源:
Ivona

例子四



應用於發音訓練的文字到視覺語音系統 (TTVS)

資料來源:
CUHK

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

語音特徵

回想一下第四課...

語音承載豐富的信息

- 含義
- 意向
- 情感
- 態度
- 個性
- 語言能力
- 口音
- 年齡
- 性別
- 健康狀況
- 教育程度
- 背景環境

我們應該在機器生成的語音中包含哪些信息？

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動一：語音特徵

收聽並描述下表中的語音特徵

語音	音量	速度	音高/音調
生氣			
長者			

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動一總結

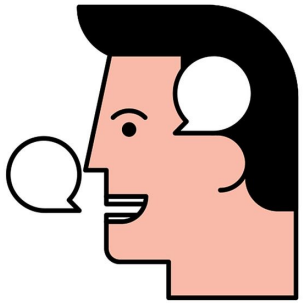
理想的TTS能於與人溝通之間產生載有豐富信息的語音

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動二：TTS的困難

在以下的錄音中，哪些是由人類說的？哪些是由機器產生的？

測試 1	A	B
測試 2	A	B
測試 3	A	B
測試 4	A	B



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

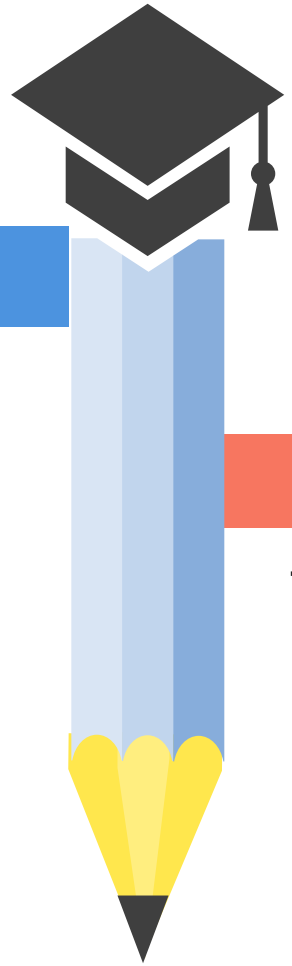
活動二的總結

1. 對於**中性語音**，TTS會產生非常**與人近似**的輸出

2. 對於**表達性的語音**，我們可以更輕易地識別出**人類與機器語音輸出的差異**

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

意識的總結



文字轉語音(TTS)的定義

文字轉語音(TTS)取**文字輸入**並產生**語音輸出**

TTS的挑戰

TTS的長期目標是能夠綜合各類的語音特徵，與聆聽者進行**有效的溝通**

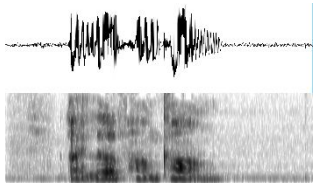
Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

知識

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

語言的階層

語音合成從**文本輸入**開始，以**合成語音波形**結束

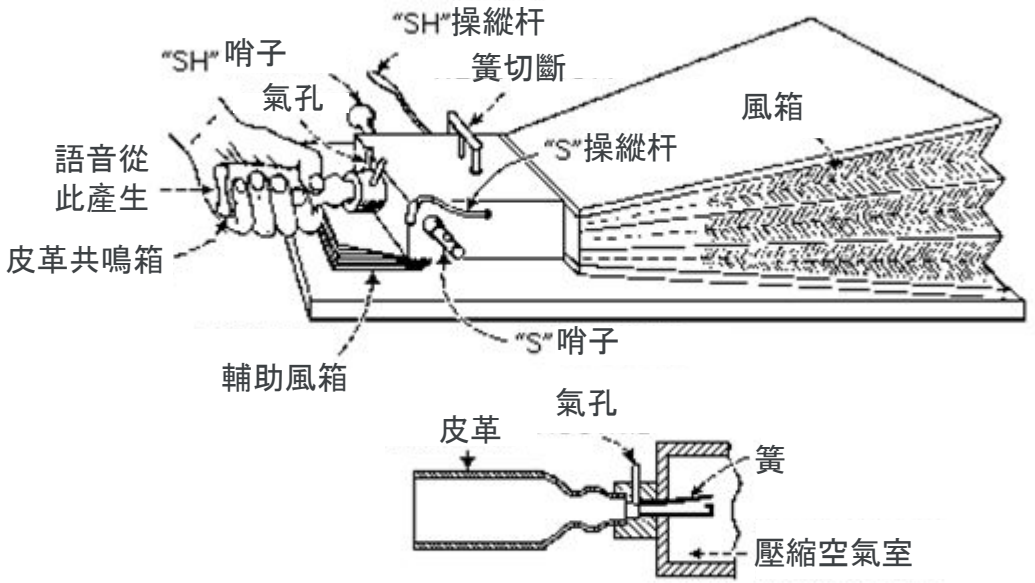
句子	***“我　愛　　香港”***
語法	I　　love　Hong Kong *代詞　動詞　名詞*
詞語 / 短語	*我* *愛* *香港*
音素	*ŋɔ ŋɔi hœŋ kɔŋ*
原聲	

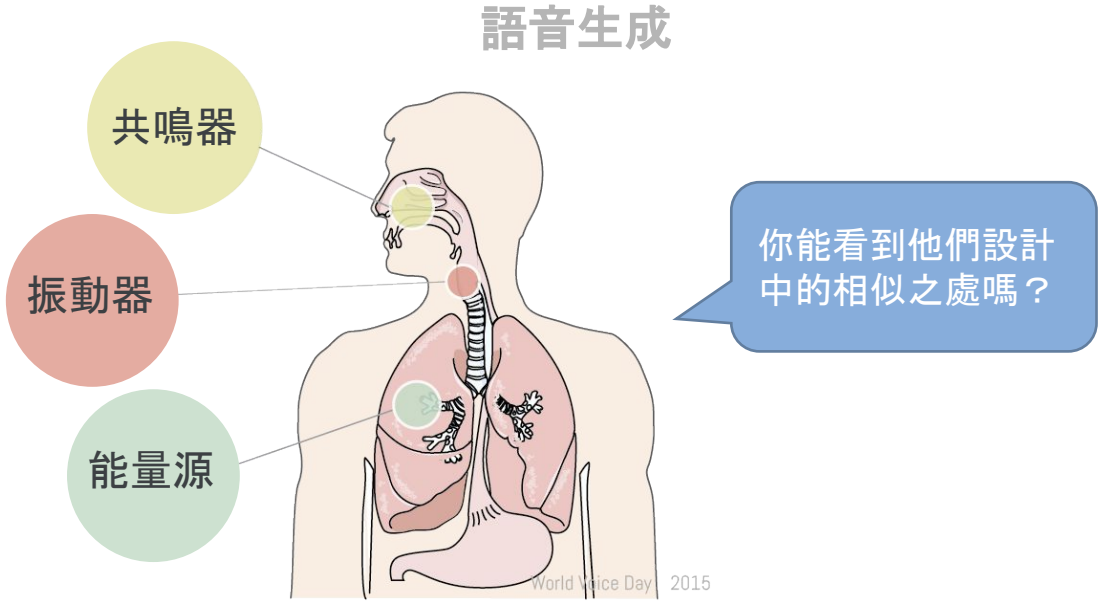
Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

歷史背景

Von Kempelen的對講機

1791年早期的設計使用氣管和發音器以模仿人類語音生成





Watch: https://www.youtube.com/watch?v=oIjkzZGe2I8

從對講機到電子系統以來，現代的語音合成和TTS都已經使用**統計算法和深度學習**來開發

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

目前的發展

如今，語音是使用**統計算法和深度學習**方法合成的

比較下面的兩個合成語音。請注意現代TTS產生的語音如何更清晰，更自然

早期TTS

現代TTS

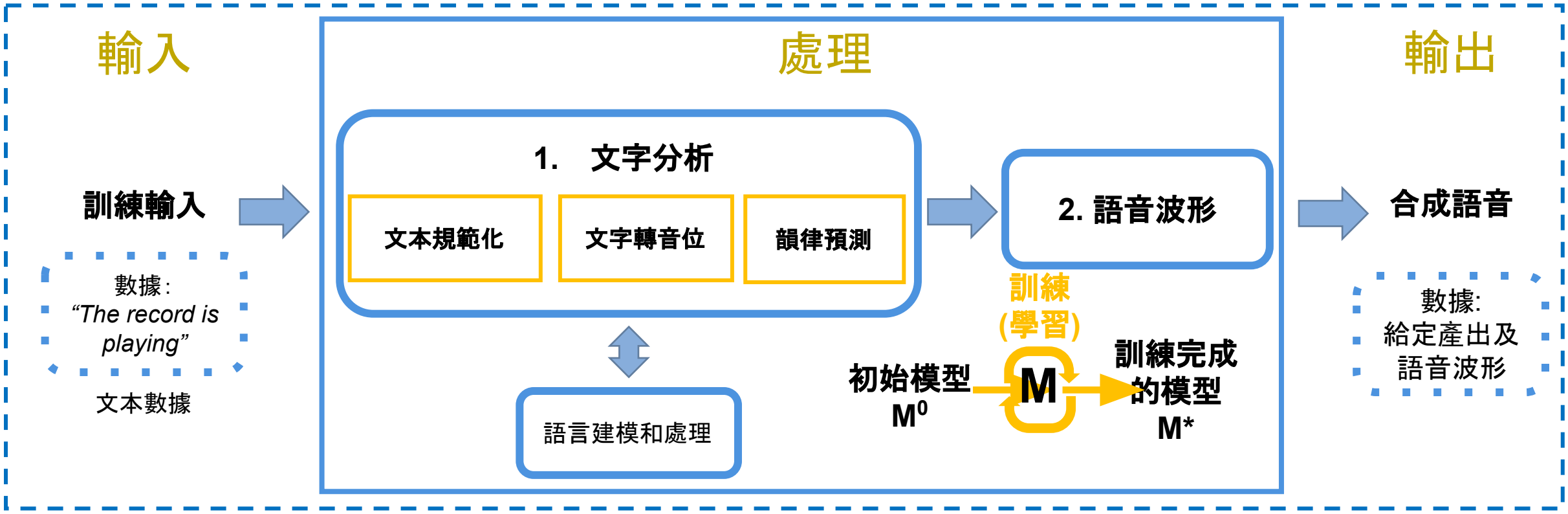
（使用大數據，雲計算，機器學習方法）

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

現代TTS方式

輸入：文本　　**輸出**：合成語音

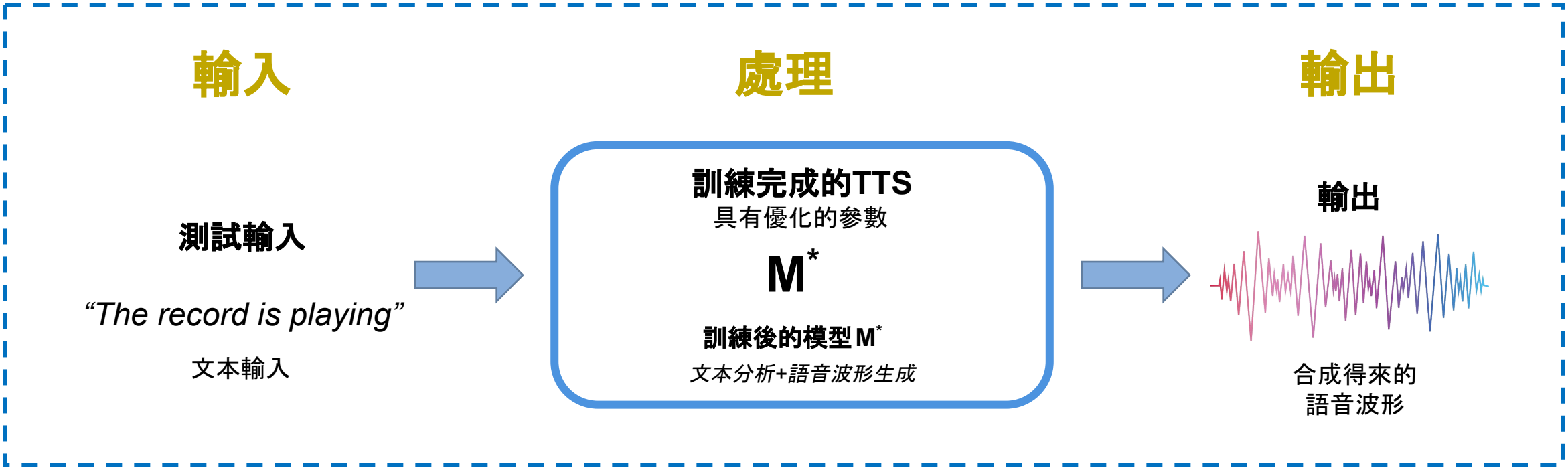
訓練過程：輸入文本和語音波形輸出合成語音，以優化內部參數



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

現代TTS方式

測試過程：基於文本的輸入，而輸出合成語音



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

1. 文本分析

此過程將**輸入文本**轉換為**符號語言形式**

文本分析的階段包括:

文字規範化

文字轉音位

韻律預測

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動一：文字規範化

請僅將以下所有語句完全轉換文字，例如

~10%➡約百分之十左右

1. “現在時間是1:25pm, 而籃球比賽的分數是3:2。”
2. “於1998, 他賣出了1998 盒巧克力, 並賺了$200 而不是 $20million。”
3. "我外號「高人」, 有成150cm高。"
4. "她弟弟出生於2014/3/31號。"

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動一總結

文本規範化是將**文本轉換為可說出的詞語**的過程

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動二：文字轉音位

請嘗試將以下文本轉換為語音**音位的標籤**

句子一: “There are many books to **read** but I have **read** them all.”

句子二: "當**單**志堅**單**獨**行**街時，看見匯豐銀**行**旁的唱片**行**正在進**行**聖誕大特賣"

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動二總結

1. 文本到音位的轉換可能會遇到多音字的問題，例如 “read”，“單”

需要進行**語言分析**以選擇正確的發音
例如：上海市長江大橋（上海市 長江大橋）（上海 市長 江大橋）
專有名詞：費先生 /bei3/ /sin1/ /saang1/
單志堅 /sin6/ /zi3/ /gin1/
詞性，例如：record: rec’ord (動詞) 對 r’ecord (名詞)

2. 文字到音位的轉換可能會使用到發音字典和字母到聲音規則

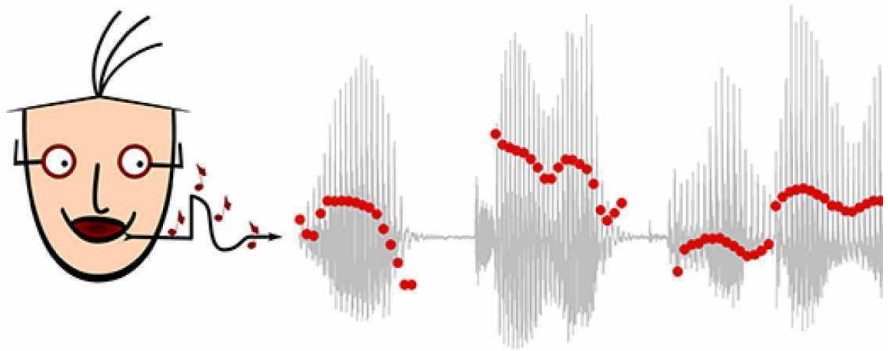
字母到聲音規則，適用於不包括於字典內的詞語，尤其適用於專有名詞或外來詞，例如“史酷比”，“Scooby Doo”

```
ABBREVIATE      AH B R IY V IY EY T
ABBREVIATED     AH B R IY V IY EY T AH D
ABBREVIATED(2)  AH B R IY V IY EY T IH D
ABBREVIATES     AH B R IY V IY EY T S
ABBREVIATING    AH B R IY V IY EY T IH NG
ABBREVIATION    AH B R IY V IY EY SH AH N
ABBREVIATIONS   AH B R IY V IY EY SH AH N Z
ABBRUZZESE      AA B R UW T S EY Z IY
ABBS            AE B Z
ABBY            AE B IY
ABCO            AE B K OW
ABCOTEK         AE B K OW T EH K
ABDALLA         AE B D AE L AH
ABDALLAH        AE B D AE L AH
```

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

韻律

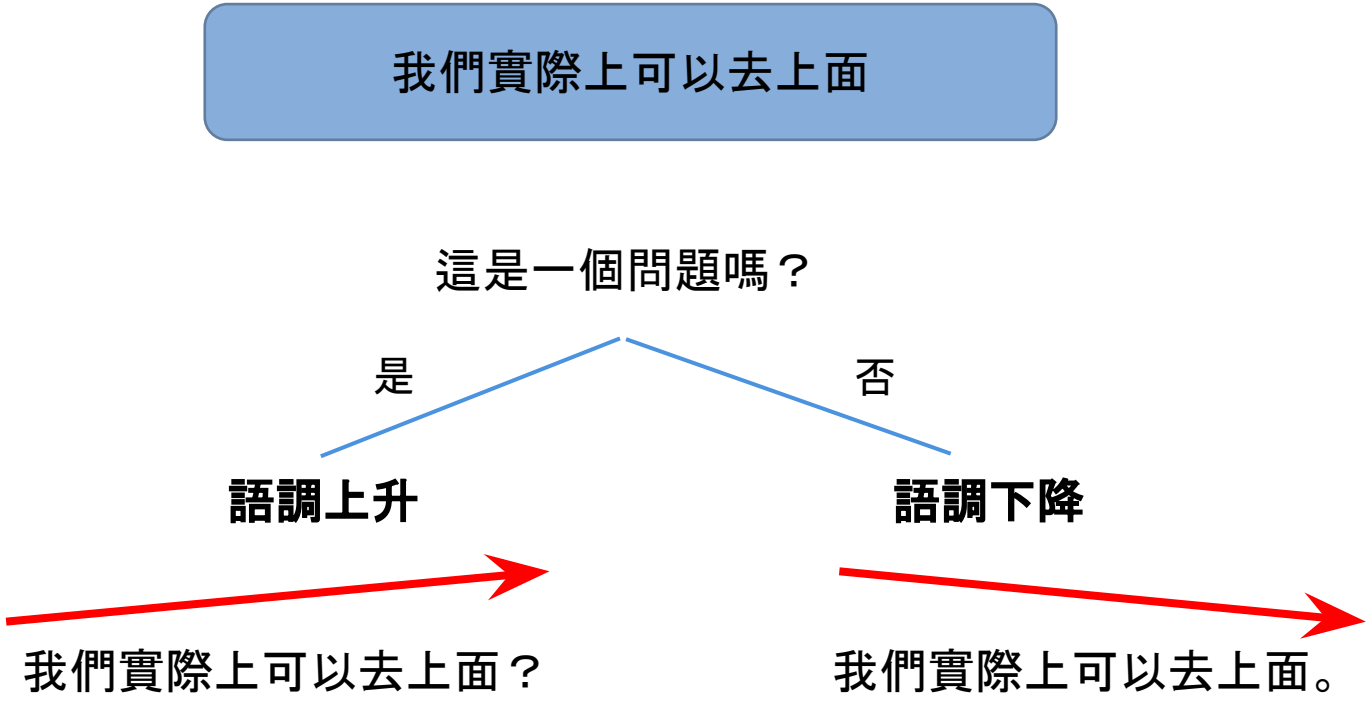
- **韻律**代表說話的方式
- **韻律特徵**包括音調，持續時間，能量(響度)等
- 韻律**有助於表達**情緒、強調以及許多不同的表達方式



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

韻律控制

韻律控制例子



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

2. 波形產生

輸入

處理

輸出

輸入文本

數據:
"The 1st record is playing"

文字分析

詞性 *限定詞* *數字* *名詞* *動詞*

The first record is playing

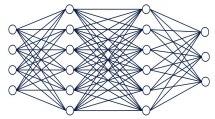
音位 *dh ax* *f uh st* *r eh k ao rd* *ih z* *p l ey y ih ng*

可以使用帶標籤的數據進行訓練

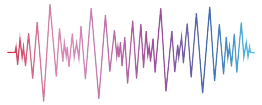
語言表示

可以使用帶標籤的數據進行訓練

音波產生



合成語音



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

TTS評估

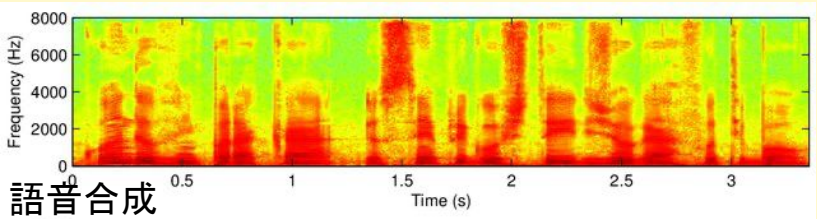
我們如何評估文字轉語音系統的表現？

客觀測量

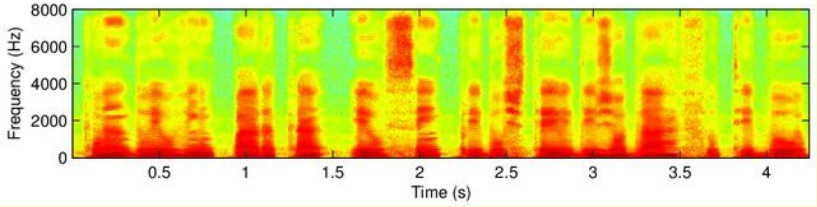
計算合成語音和人類自然生成語音之間的語音波形相似度

例如：信號距離

自然語音



語音合成



主觀測量

一群人評價合成語音的質量

例如：平均意見分數

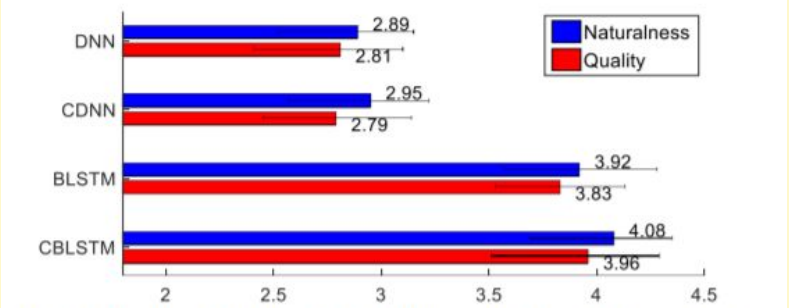


Fig.4. Results of MOS test with 95% confidence intervals for speech naturalness and quality.

[Li et al., 2018]

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動三：TTS評估

主觀測量

- 聆聽下面的4個合成語音，並根據語音的“可理解性”，“自然性”和“質量”以5分制來評分（1=極差，5=極好）
- 收集其他同學的一些評價，並計算4個音頻文件的平均分數

語音 1

語音 2

語音 3

語音 4

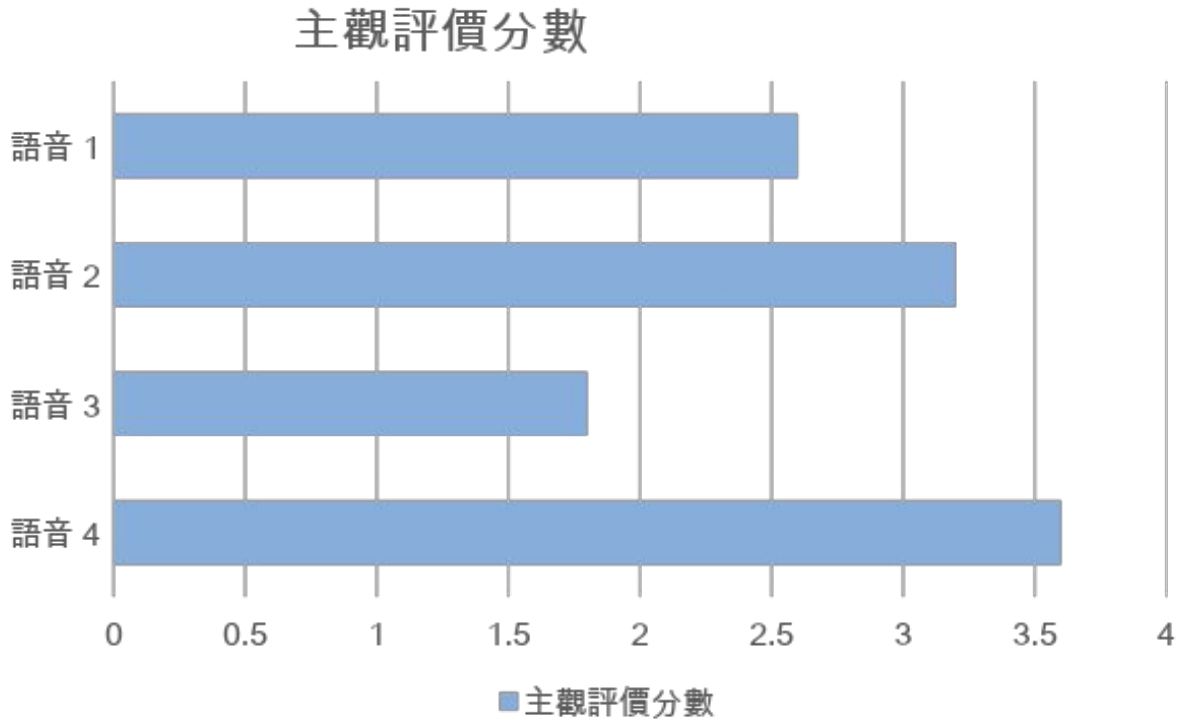
Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動三總結

我們從一組聽眾那裡收集評分分數，並獲得平均評分作為主觀評價(Subjective Evaluation)
例如於下面的示例：

評估示例:

語音 1	3, 2, 2, 3, 3 | 平均 = 2.6
語音 2	4, 3, 3, 3, 3 |平均= 3.2
語音 3	1, 2, 2, 2, 3 |平均= 1.8
語音 4	4, 4, 3, 3, 4 |平均= 3.6



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

TTS的擴展應用

語音轉換

將TTS合成語音轉換為目標說話者的語音

	原本	語音轉換	目標
1			
2			

範例：香港中文大學的技術

https://sites.google.com/site/2016icme/

視覺語音合成

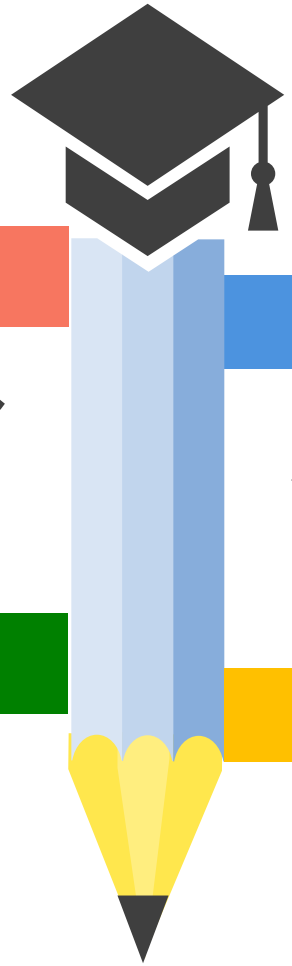
也稱為頭部特寫或頭像

專注於同步說話頭的嘴唇和語音

展示
Demo

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

知識的總結



語音合成技術

現代語音合成技術的開發**需要大數據以進行訓練**

文字分析

分析文本以進行**文本規範化，文字轉音位和韻律預測**以產生可說出詞語

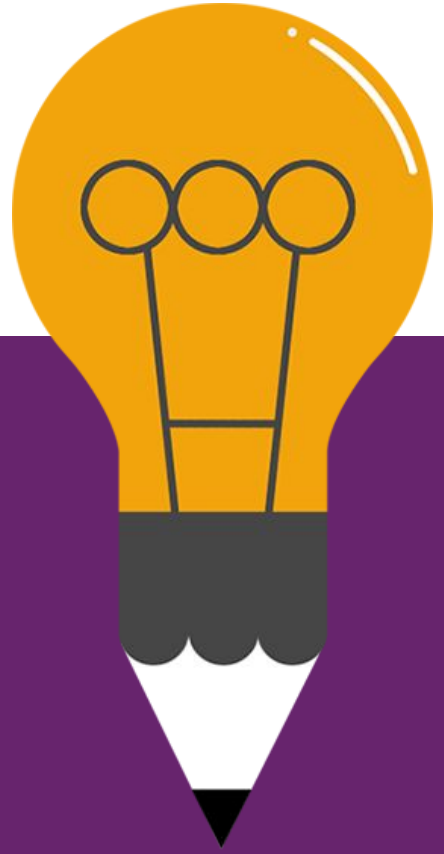
評估

TTS的評估包括客觀和主觀評估

語音波形產生

通過使用標籤了的大數據進行訓練，將語言與語音對應

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.



倫理

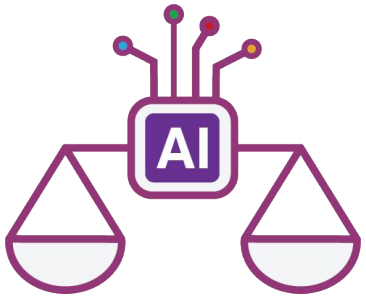
Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動一：公平問題

當你探索各種TTS展示時，為什麼它們通常不提供廣東話？

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

公平問題



Locale	Language	Gender
de-DE	German (Germany)	Female
en-US	English (US)	Male
en-US	English (US)	Female
it-IT	Italian (Italy)	Female
zh-CN	Chinese (Mainland)	Female

TTS必需有**大數據**來進行訓練

數據不足的語言將無法得到語音技術的良好支持

例如，目前微軟神經語音在當前版本中只支持四種語言

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

活動二: 安全性問題

請觀看視頻並考慮:

如果使用此技術生成的仿造聲音,
會出現哪些安全問題?

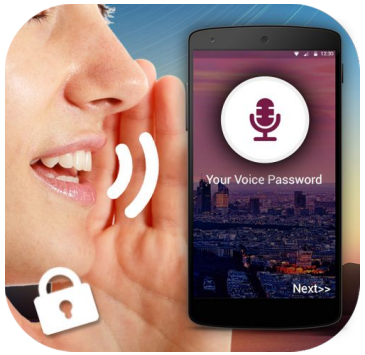
Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

安全性問題

“一個人的聲音也是他身份的一部分”

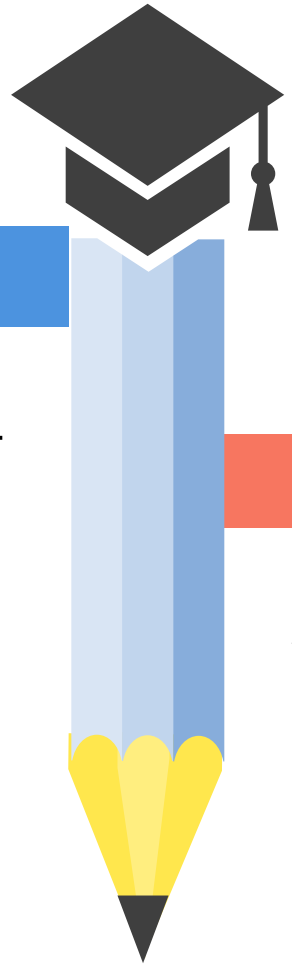
語音模擬攻擊

例如：電話詐騙和對語音生物識別系統的攻擊



Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

倫理的總結



公平問題

由於缺少帶註釋的訓練數據，TTS無法很好地支持少數群體中資源匱乏的語言

安全性問題

若用作身份盜用，個性化的仿造語音會變得既危險又有害

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.

謝謝

第五課：說

Copyright ©2019-2022 CUHK Jockey Club AI for the Future Project. All rights reserved.