SYNOPSYS®

2023 Synopsys ARC AIoT Design Contest

好好說話 The Art of Saying

Topic - 聲紋打卡系統

Presenters - 徐振逢、陳晶、楊永琪、吳維誠

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- Project Detail
- Results
- Overall Summary

Agenda

- **Introduction**
  - 作品設計動機
  - 預期應用場景與功能
- Contribution
- Challenge and Innovation
- Design and Reliability
- Project Detail
- Results
- Overall Summary

Introduction - 作品設計動機

- 新冠肺炎 COVID-19
  - 後疫情時代；不適合施打疫苗；諸多變種，打過疫苗也無法完全免疫。
  - 這次疫情，為人類生活帶來永久的改變。
    ➢ 因此，想要打造一個零接觸的打卡系統、認證方式。
  - 具唯一性且足夠方便，映入眼前的就是生物特徵辨識 — 聲紋。
  - 聲紋辨識 (Voiceprint Recognition) 比起指紋、臉部等識別方式更具衛生、便利性，只需透過聲音，無需透過身體接觸或作特定動作，即可進行辨識。
- 身分冒充問題
  - 因使用生物特徵辨識，難以冒充，有助於避免「代打卡」等問題。

Introduction - 作品設計動機

- 環境友善
  - 不需印製打卡用識別證，有利於環保。
- 職場氣氛
  - 員工進入辦公室時，與同事打招呼，增加工作職場朝氣，與此同時，系統接收訊號完成打卡，節省打卡手續。如上班打卡用「早安」等。
- ARC EM9D AIoT DK 開發板
  - 開發板尺寸輕便與 AI 模型運算之硬體優化，有助於關鍵字識別 (Keyword Spotting, KWS) 模型的部署和運行速度。
  - 語音系統在商業環境並未普及，期望利用 ARC EM9D 開發板，讓生活更方便、更友善。

Introduction - 預期應用場景與功能

- 商用打卡系統：幫助員工快速、便利地打卡。
- 課堂點名：簡化教師點名程序。
- 門禁系統：聲紋作為生物特徵具有唯一性，與身份權限結合控制存取權，增加安全性。
- 汽機車引擎發動：除了遙控鑰匙，另一項選擇或是兩項都採用增加安全性。
- 無卡提款、電子護照、電腦解鎖：與身分驗證、機密資料存取等相關操作。

Agenda

- Introduction
- **Contribution**
- Challenge and Innovation
- Design and Reliability
- Project Detail
- Results
- Overall Summary

Contribution - Optimization Technology

- 優化方法
  - 量化模型、減少模型層數及參數量：減少模型記憶體用量、加快運算速度
  - 使用 Streaming model：減少重複計算、增加真實情境準確性
  - 使用 MLI library：加速推論速度
  - 優化移植的運算子：減少推論計算量
  - 平行處理推論與收音：減少無收音的情況，增加裝置執行效率

Agenda

- Introduction
- Contribution
- **Challenge and Innovation**
- Design and Reliability
- Project Detail
- Results
- Overall Summary

Challenge

- Model Size
  - RAM size ~ 2MB
  - 若將模型縮小，會使得模型的準確度下降、難以訓練。
  - ➔ 1.4 MB 縮小為 183 kB
- TFLM Operators NOT support
  - MFCC Op
  - Audio Spectrogram Op
  - SUM Op
  - ➔ 移植成功，並優化效率

Challenge

- Massive Computation
  - 為了更精準的捕捉關鍵詞，在擷取輸入音訊時，會有所重疊。
  - 導致推論時，需進行大量的重複計算。
  - 對於計算資源有限的微處理器架構的物聯網裝置而言
    - 這拖累系統性能，甚至使其無法正常運行。

➔ 導入 Streaming 機制，消除重複計算，降低延遲。

Challenge

- Information Loss
  - 關鍵詞辨識模型之輸入為 1 秒的音訊，在 ARC EM9D 推論所需的時間為 1.35 秒
    - 一次推論需 2.35 秒，且其中有大量時間無法收音的困境。
  - 資訊流失：~60%
  - ➔ 優化：平行處理麥克風與推論

Challenge

- High Latency
  - Audio Spectrogram 與 MFCC：占總延遲的 95.8%。
  - DSP Library Not support
    - 讓硬體支援的傅立葉轉換 (RDFT) 與三角函數等運算。
- Limited Memory
  - 由於 ICCM1 只提供 320 KB 的容量，在整合階段時會面臨記憶體不足的問題。

Innovation

- 將聲紋辨識系統引入商用環境。
- 將語音辨識及其相關運算子引入 ARC EM9D AIoT DK 開發板。
- Streaming 機制：減少模型推論延遲。

Innovation

- 將聲紋打卡系統分為兩個部分：關鍵詞偵測、聲紋辨識。

  改良傳統單一模型的做法：

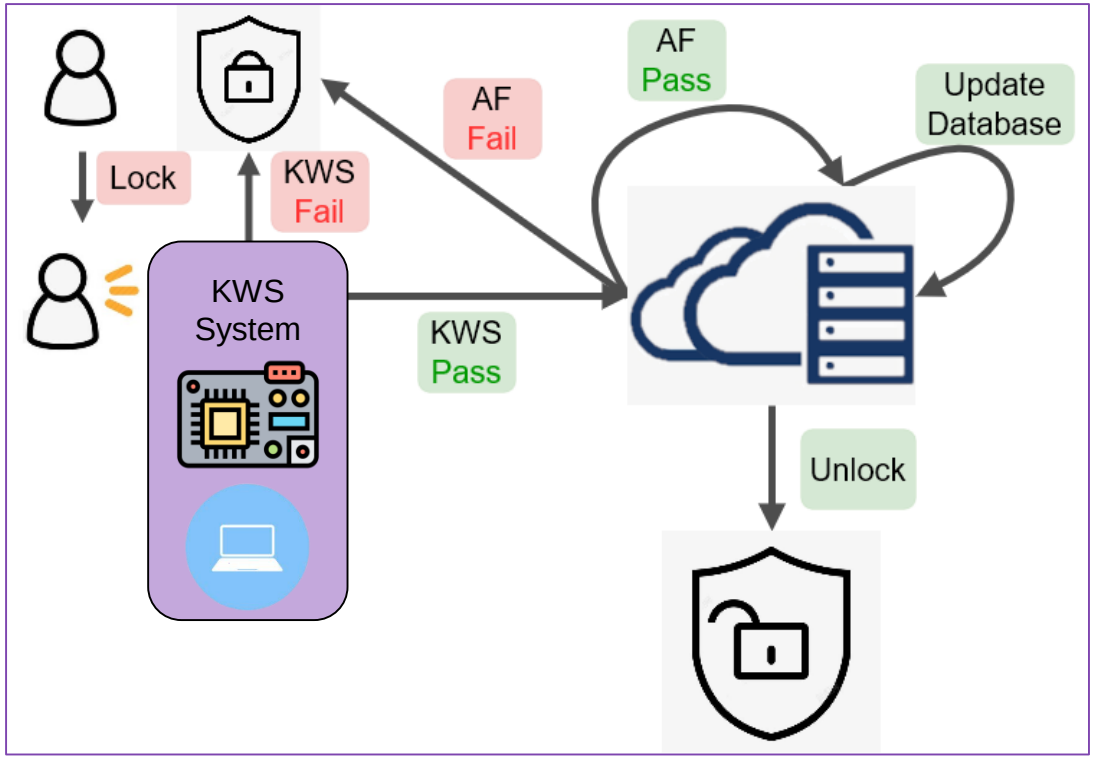
  - 門禁的聲音偵測裝置(物聯網裝置)：負責關鍵詞偵測，無需維持龐大的聲紋辨識模型
  - 雲端：負責聲紋辨識
    - ➔ 節省待機與運算時的能源消耗
    - ➔ 有益於系統的維護與管理
    - ➔ 減輕邊緣裝置的負擔，使其更加輕便、快速

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- **Design and Reliability**
- Project Detail
- Results
- Overall Summary

Design and Reliability

- 設計與流程：
  1. 利用 Streaming KWS 模型偵測關鍵詞。
  2. 將該音訊傳至雲端(如 GCP、AWS、公司內部伺服器...等)，進行聲紋辨識。
  3. 當說話者成功驗證身份後：
     - 回傳結果並進行解鎖
     - 雲端寄送通知
     - 在雲端資料庫中更新出缺勤情況，確保機密資料被安全地存儲於雲端



Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - **Keyword Spotting Model**
  - KWS on ARC EM9D
  - Voiceprint Recognition Model
  - Database
  - Connection between ARC EM9D and Cloud
- Results
- Overall Summary

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - **Keyword Spotting Model**
    - **Dataset**
    - Model Selection and Improvement
    - Streaming Model
- Results
- Overall Summary

Project Detail - Dataset

- 訓練資料集：
  - TensorFlow 的 speech_commands_v0.02 資料集
  - 內含 35 種詞彙
  - 共有 105,829 筆資料。
- 採用「**happy**」為本專案關鍵詞，其餘皆歸類為 「unknown」。
- 每筆資料為 1 秒 16-bit、16000 Hz 的 PCM 錄音檔。

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - **Keyword Spotting Model**
    - Dataset
    - **Model Selection and Improvement**
    - Streaming Model
- Results
- Overall Summary

Project Detail - Model Selection and Improvement

- 為使關鍵詞辨識模型有更好的整體表現，採取以下作法：
  1. 使用 Singular Value Decomposition Filter (SVDF) ➔ 降低模型計算量
  2. 將模型的權重從 32 位元浮點數量化成 8 位元整數 ➔ 降低模型大小
  3. 減少模型層數與參數量
  4. 採用 Google KWS Streaming model 之架構 ➔ 加快推論速度、減少重複計算

Project Detail - Model Selection and Improvement

- Float32 Non-Stream 模型測試結果

	Model size (KB)	Accuracy	Int8 Model Size	Int8 Accuracy	Latency (ms)
Original SVDF	1.4MB	96.54	418	91.04	x
SVDF 4 Layers	834	95.56	256	91.72	1275*
SVDF 2 Layers	335	93.72	100	86.73	1373
SVDF 2 Layers - 2	183	93.29	58	83.87	1349

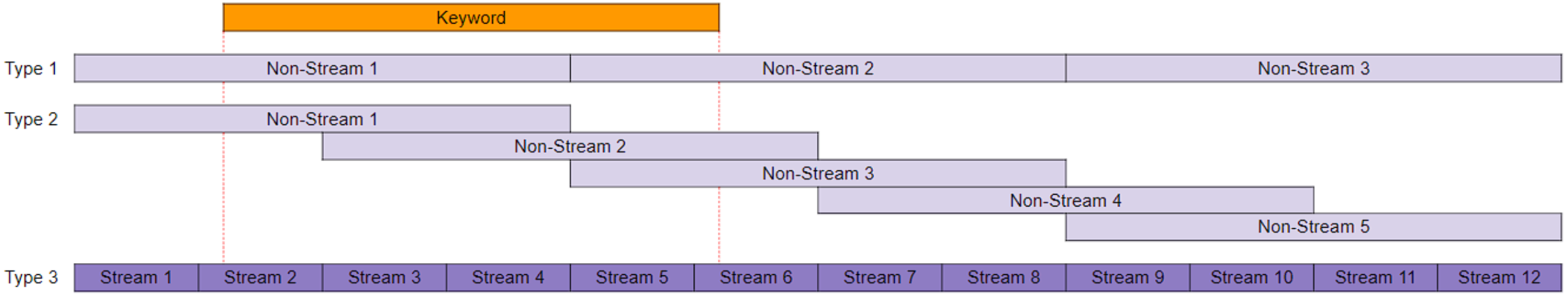
*使用 Int8 模型測量

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - **Keyword Spotting Model**
    - Dataset
    - Model Selection and Improvement
    - **Streaming Model**
- Results
- Overall Summary

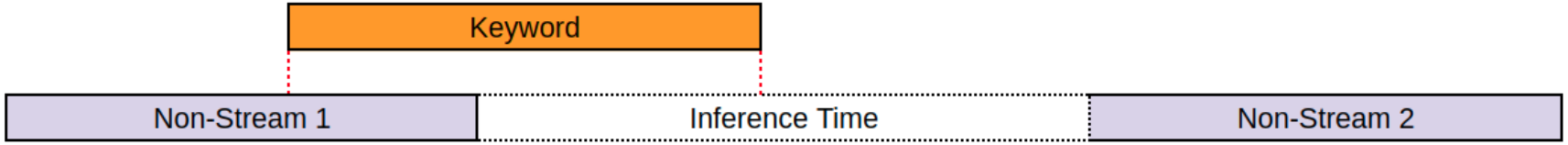
Project Detail - Streaming Model

- 動機
  - 在傳統的關鍵字辨識模型中，有大量的重複計算，增加運算成本與時間。



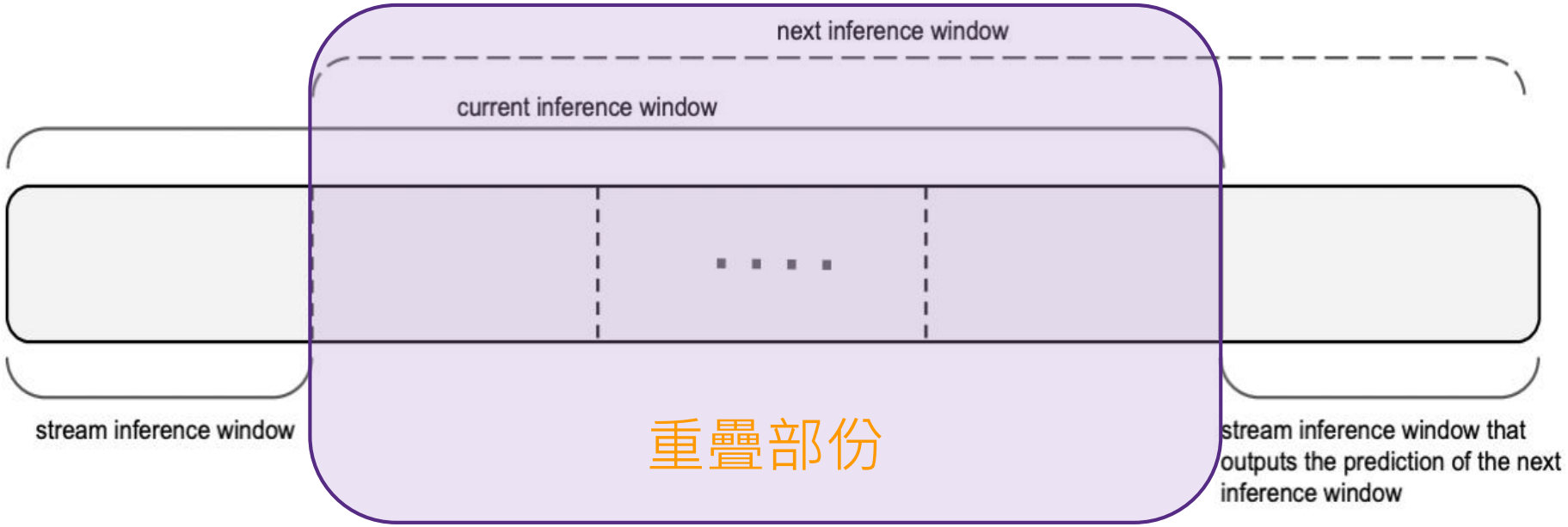
Project Detail - Streaming Model

- 動機
  - 對於計算資源有限的單核心架構的物聯網裝置而言，上述缺點更顯得不容忽視，可能拖累系統性能，甚至使其無法正常運行。
    - 因 ARC EM9D 推論所需時間甚巨，導致每擷取 1 秒的音訊後，還需等待 1.35 秒的困境。
    - 2.35 秒只完成一次推論，其中大量時間無法接收聲音，難以接受。
    - 資訊流失：57.4%。
    - 實際測試於 ARC EM9D 開發板上，無法正常進行關鍵詞偵測。



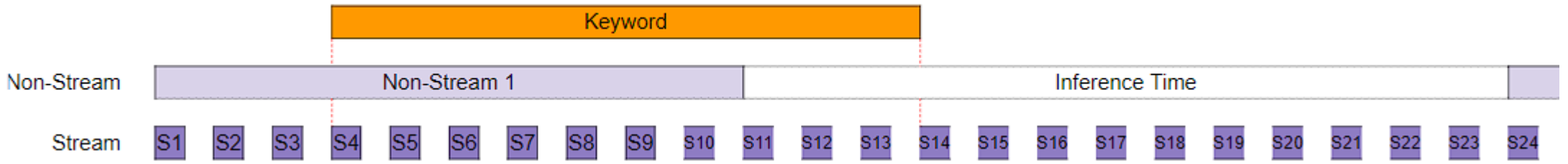
Project Detail - Streaming Model

- 方法
  - 保存當前訊框的部分計算結果，供下一個訊框推論時使用，而省去兩者之重疊部分所導致的重複計算。



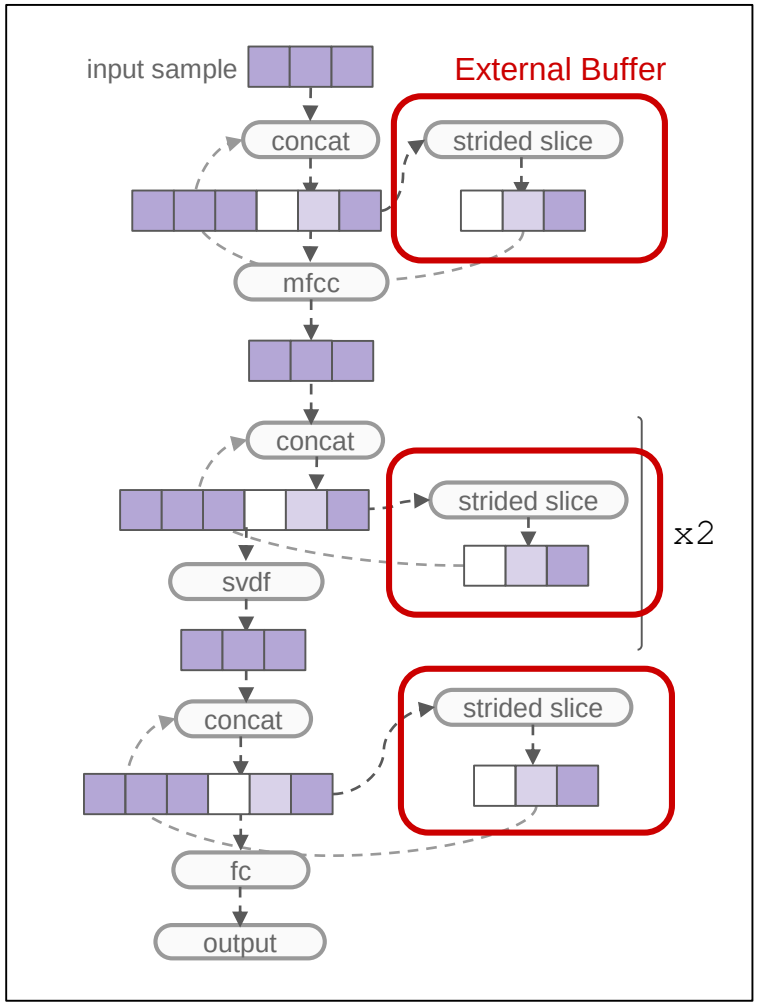
Project Detail - Streaming Model

- 運作方式
  - 減少輸入音訊長度 ➡ 提昇推論效率、減少延遲
  - 利用短時間收音，短時間推論的方式 ➡ 更均勻的採集聲音
  - 透過暫存上一次運算的結果，與當前輸入值一起計算。
    - 讓聲音特徵(音訊中斷)損失最小化、成本最小化。
    - 與一般的 Non-Streaming model 達到相同效果。



Project Detail - Streaming Model

- 採用 SVDF 2 Layers - 2 External Streaming Model 。
- 模型擁有 Multi-Input/Ouput 的構造，使用 External Buffer 暫存重疊計算的部份，作為下一次推論的輸入。
- 輸入音訊長短：
  - Non-Streaming model：1s
  - Streaming model：20 ms
- 推論延遲：
  - Non-Streaming latency : 1349 ms
  - Streaming latency : 31 ms



Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - Keyword Spotting Model
  - **KWS on ARC EM9D**
  - Voiceprint Recognition Model
  - Database
  - Connection between ARC EM9D and Cloud
- Results
- Overall Summary

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - **KWS on ARC EM9D**
    - **Port TFLite op to TFLM**
    - Integration with Mic and LED
- Results
- Overall Summary

Project Detail - Port TFLite op to TFLM

- 將 Audio Spectrogram、MFCC、SUM 運算子加入 TFLM
  - 從 TensorFlow Lite 將運算子移植到 TFLM。
  - 優化：
    - 在運算資源不充足的情況下，將動態記憶體配置之資料結構 (如：vector、dequeue 等)，修改為靜態記憶體配置，以利效能最佳化。
    - 將運算子內初始化的動作從 TFLM 中 Eval 階段調整到 Prepare 階段，並將記憶體宣告置於 PersistentBuffer 內，使得在 Prepare 階段的計算之結果得以保存，避免每次推論時的重複計算。
    - 結合兩項優化減少 121ms 的推論延遲，即 79.6% 的 Streaming model 延遲、7.8% 的 Non-Streaming model 延遲。

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - **KWS on ARC EM9D**
    - Port TFLite op to TFLM
    - **Integration with Mic and LED**
- Results
- Overall Summary

Project Detail - Integration with Mic and LED

- 開發板的麥克風與模型的即時推論
  1. 整合音訊擷取的函式庫 (aud_lib)。
  2. 優化：將收音之運作模式，從 **Busy Waiting** 等待，改善為收音與推論同時進行。

```
res [[silence]], cnt 63
aud_rbf_of w_idx: 0
res [[silence]], cnt 64
res [[silence]], cnt 65
res [[silence]], cnt 66
aud_rbf_of w_idx: 0
res [up], cnt 6
res [up], cnt 7
res [up], cnt 8
aud_rbf_of w_idx: 0
res [up], cnt 9
res [up], cnt 10
res [up], cnt 11
aud_rbf_of w_idx: 0
aud_rbf_of w_idx: 0
res [left], cnt 6
res [left], cnt 7
res [left], cnt 8
res [left], cnt 9
aud_rbf_of w_idx: 0
res [left], cnt 10
res [left], cnt 11
res [left], cnt 12
```

Log on ARC EM9D

Project Detail - Integration with Mic and LED

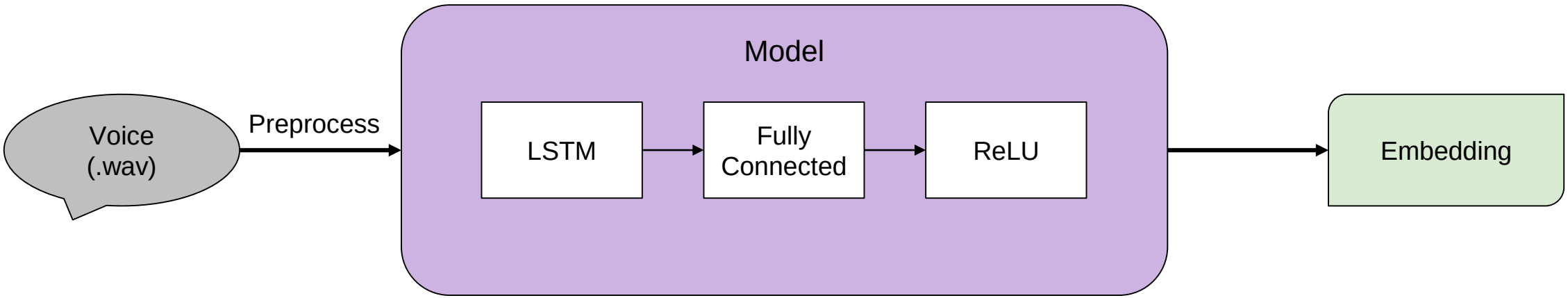
3. 將 GPIO 函式庫 (hx_drv_iomux) 控制 LED 燈，作為系統不同階段之提示燈：
   - 紅色燈號：系統正常運行。
   - 藍色燈號：已偵測到關鍵詞，並將音訊傳送至雲端，等待聲紋辨識結果。
   - 綠色燈號：聲紋辨識結果為「PASS」，通過驗證。
   - 當藍色燈號熄滅，且綠色燈號沒亮：表示聲紋辨識結果為「FAIL」，驗證失敗或沒有權限。

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - Keyword Spotting Model
  - KWS on ARC EM9D
  - **Voiceprint Recognition Model**
  - Database
  - Connection between ARC EM9D and Cloud
- Results
- Overall Summary

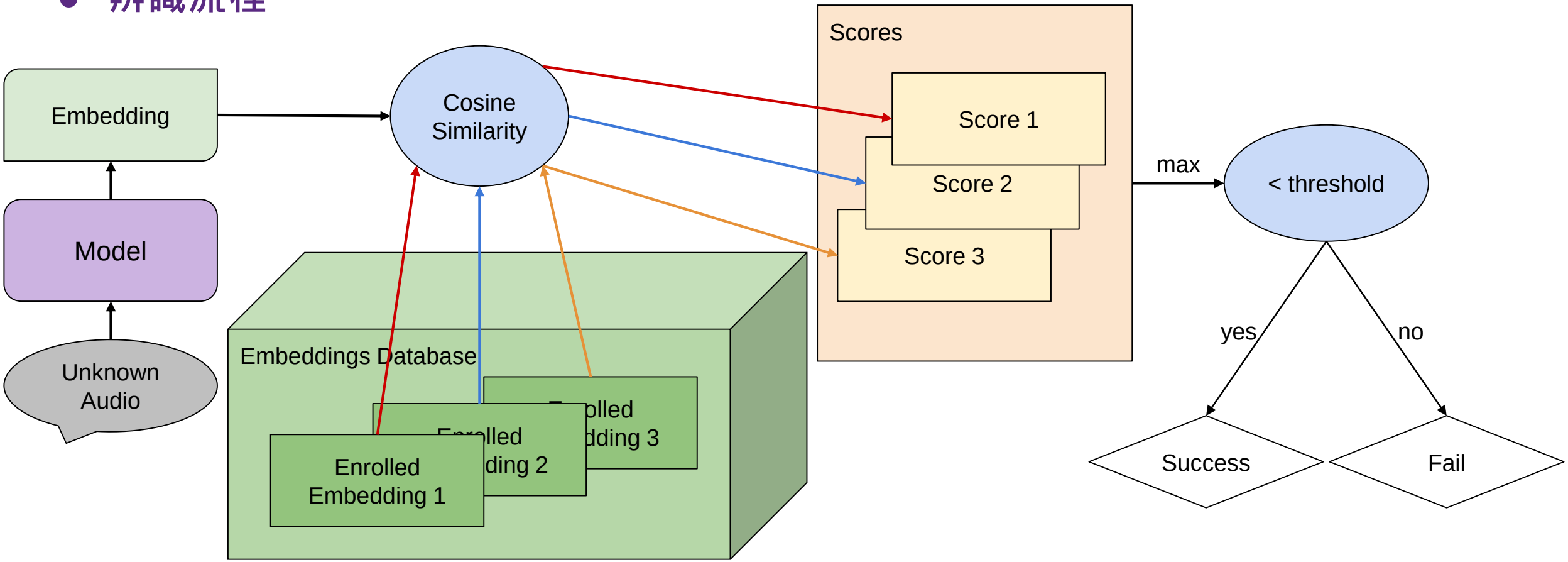
Project Detail - Voiceprint Recognition Model

- 模型設計



Project Detail - Voiceprint Recognition Model

- 辨識流程

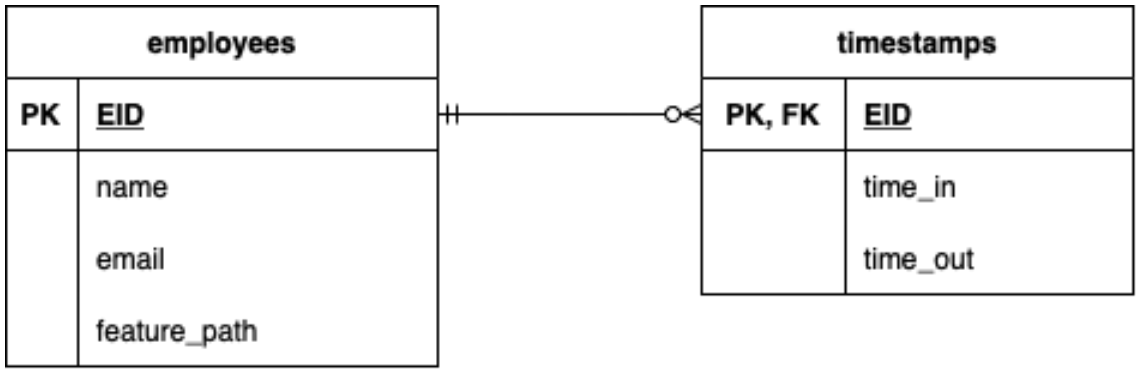


Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - Keyword Spotting Model
  - KWS on ARC EM9D
  - Voiceprint Recognition Model
  - **Database**
  - Connection between ARC EM9D and Cloud
- Results
- Overall Summary

Project Detail - Database

- **資料庫設計**
- ER Diagram：



  - Employees：職員ID、職員姓名、職員電子信箱、職員聲紋 embeddings 之檔案位置
  - Timestamps：職員ID、打卡時間（上班）、打卡時間（下班）
- 實作：SQLite

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- **Project Detail**
  - Keyword Spotting Model
  - KWS on ARC EM9D
  - Voiceprint Recognition Model
  - Database
  - **Connection between ARC EM9D and Cloud**
- Results
- Overall Summary

Project Detail - Connection between ARC EM9D and Cloud

- **Serial Port**

- 目前使用序列埠作為 ARC EM9D 開發板，與外部連接的界面。
- 需額外的裝置（電腦）作為中繼站：
    1. 從 ARC EM9D 開發板，接收音訊資料
    2. 轉換成 wav 檔
    3. 傳送至雲端
    4. 接收結果
    5. 回傳給 ARC EM9D 開發板
- 從 ARC EM9D 開發板傳輸「長度 1 秒的音檔」到中繼站需耗時 8 秒。

Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- Project Detail
- **Results**
- Overall Summary

Results

- 最終成功部署 Streaming 模型，達到即便計算資源有限，仍能即時辨識出關鍵詞。
  - 與原先 Non-Streaming 模型相比：
    - ➔ 單次推論延遲時間變成：原本的 2 ~ 3%
    - ➔ 將收音比率從 42% 提高至約 70%
    - ➔ 模型大小：1.4MB 縮小為 183kB (13%)

- 將網路傳輸由 Wi-Fi 改為序列埠：
  - ➔ 將網路傳輸時間從約 20 秒縮短至 8 秒 (40%)

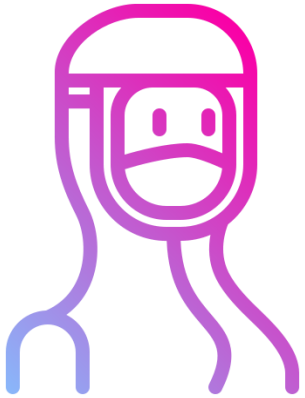
Agenda

- Introduction
- Contribution
- Challenge and Innovation
- Design and Reliability
- Project Detail
- Results
- **Overall Summary**

Overall Summary

portable

always-on



contactless

clock-in system

Thank You

SYNOPSYS®