

混合語言之語音的語言辨識

Author : 朱晴蕾 呂道誠 呂仁園

Professor : 陳嘉平

Reporter : 許峰閣

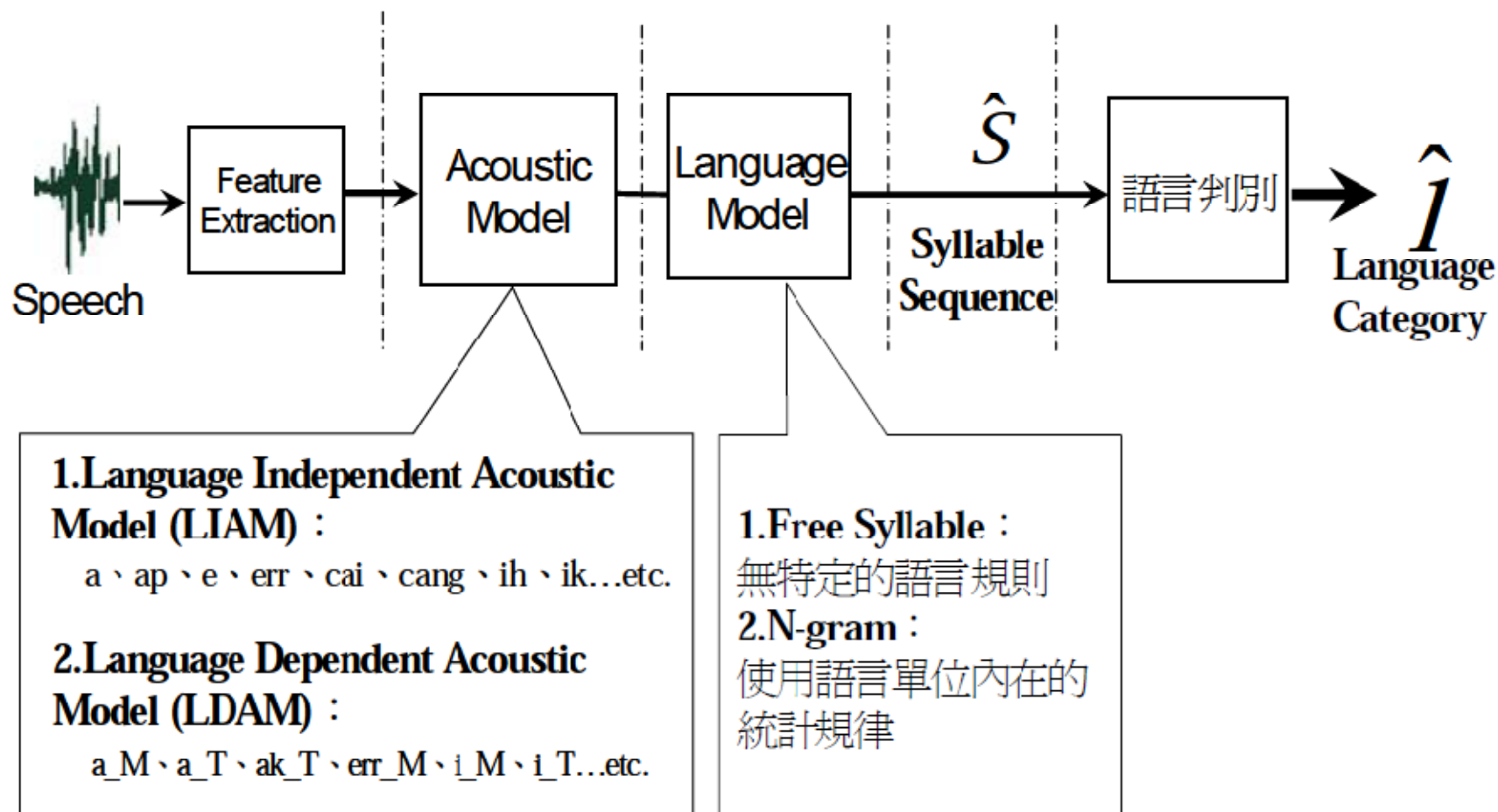
大綱

- 緒論
- 音節辨識
- 語言判別
- 實驗

緒論

- 傳統的**LID**主要是用於判別出一段語音是何種語言所構成,但是現在語言轉換(**Code-Switching**)的技巧在日常生活中平凡的出現
- 所謂**Code-Switching**就是兩種或兩種語言以上交體組合而成的語言
- 此篇**PAPER**辨識希望能區分一段語音中部同語言的交換分界處

系統架構圖



圖二、系統架構流程圖

音節辨識

- 在語言的標音上,音標使用**ForPA**,在華語的音素有**37**個,台語的音素有**56**個,聯集有**63**個,交集有**32**個
- 分爲聲學模型跟語言模型兩部分
- 在聲學模型方面,分成語言獨立的聲學模型以及語言相依的聲學模型

聲學模型

- 語言獨立的聲學模型: 聲學模型中所有的音標沒有語言上的分別,是由華語及台與共同訓練而成
- 語言相依的聲學模型:針對每種語言分別訓練其獨立的聲學模型,並在每個模型中標記其所屬的語言類別,再結合在一起,所以其音標都有所屬語言的標籤

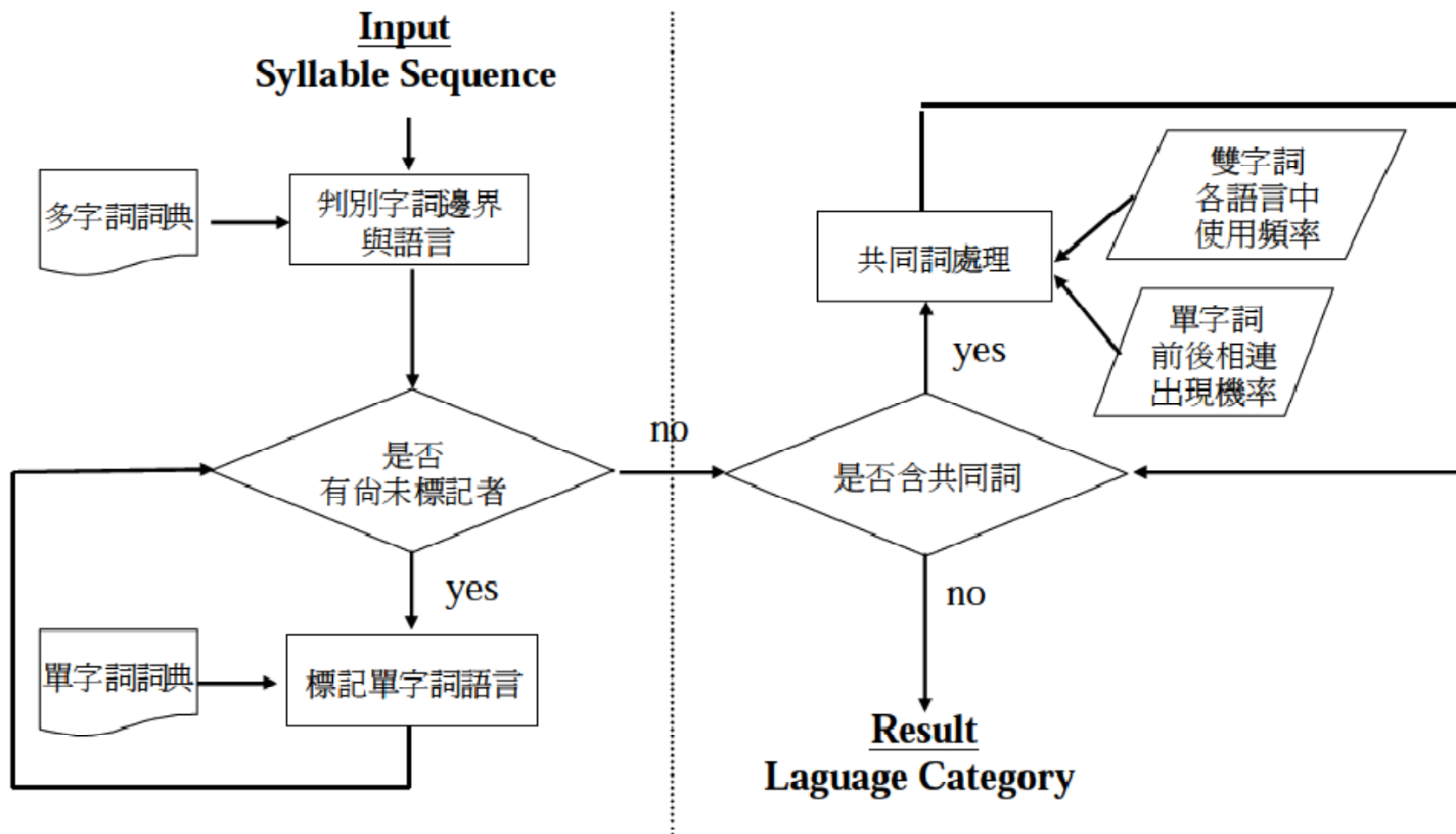
語言模型

- **N-gram**的語言模型可以提供一種語言中其文字的序列規則,並以統計和機率的方式來呈現
- 對正確的語言現象,字與字之間共同出現機率較高,對較不符合語法者,字與字之間共同出現機率較低,例如:‘今日’ 這種組合在華語中機率較高,‘今仔日’ 在台語中較高

語言模型

- 在這邊是使用**bi-gram**的語言模型,並以音節為單位量,再用約一萬七千句華語及九千句台語來訓練這個模型,所有句子皆僅由一種語言組成

語研判別



圖三、語言判別流程圖

判別字詞邊界與語言

- 在多字詞詞典方面,使用中研院的華語詞典及實驗室中的台語詞典混合,總詞數約十四萬詞,並將每個多字詞後標記所屬語言種類,華語約八萬詞,台語約五萬詞,兩種語言皆可能出現者一千五百詞
- 單字詞字典華語單字詞數約**230**個,台語為**580**個,華語台語共同出現為**200**個

判別字詞邊界與語言

- 在實際觀察混合語言的語音文句中,發現語言的轉換時機都在詞(**word**)上,在判別字詞邊界與語言的這個步驟,變是將音節序列與詞典中詞句做最長匹配比對,將比對到的部分做爲詞與詞的分界點
- 例如:我最喜歡去(夜市仔)吃東西便可以切成我最喜歡去、夜市仔、吃東西 這三個詞
- 但音節組合可能不在詞典中,所以使用單字詞詞典來解決這個問題

共同詞處理

- 共同詞處理又分爲多字詞及單字詞,在多字詞的共同詞處理中,使用統計語料庫出現的頻率高低來決定,取出現機率較高者
- 單字詞的共同字處理分爲兩種
 - 1.語言獨立聲學模型(LIAM)
 - 2.語言相依聲學模型(LDAM)

共同詞處理

- 在**LIAM**的模型下,每個音節都會帶有語言標籤,便直接以該語言標籤為依據
- 在**LDAM**的模型下,在做語言轉換的時候估計都是以**word**為邊界,所以若前後兩個音節為同一語言時,該音節即與前後語言為同語言,但此法並無法完全解決問題,所以現多採用**N-gram**的概念,利用前後音節同時出現的機率還決定該語言

實驗

- 在混合語言的文句上,收集了**75**句,由十個人錄音共**750**句,在安靜無噪音的環境下錄製**16kHz,16bit**的聲音訊號

表一、華台語混合語言語音例句

| Filename | Text | Transcription |
|----------|---------------|---|
| MT_001 | 我最喜歡去（夜市仔）吃東西 | uo3_zuei4_si3_huan1_cyu4_ia2_ci2_a4_chii 1_dong1_si1 |
| MT_005 | （歹勢！）我遲到了 | painn4_se3_uo3_chii2_dau4_ler0 |

實驗

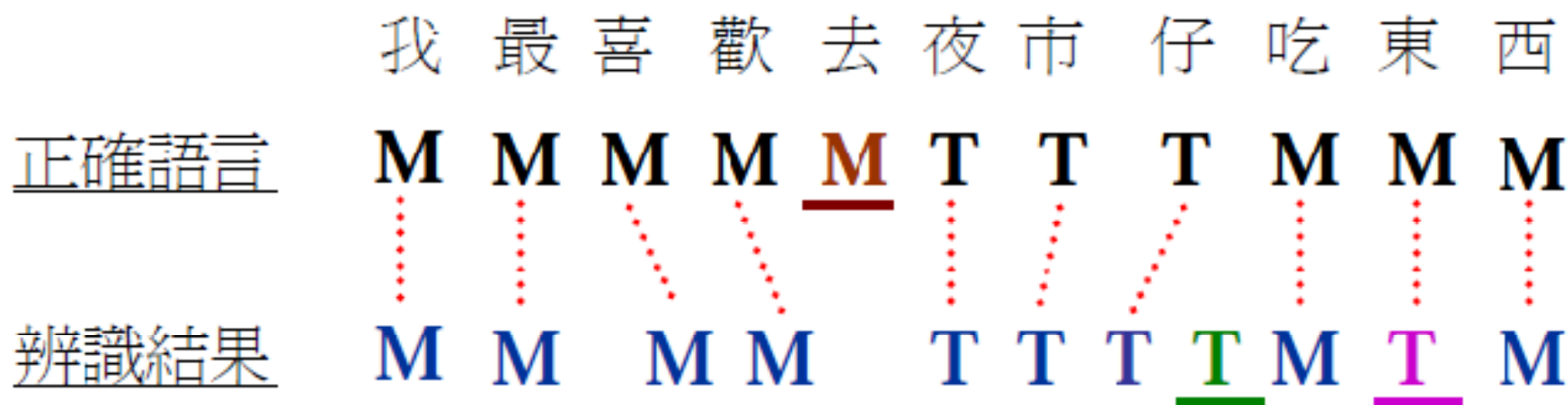
- 使用兩種不同的聲學模型及兩種不同的語言模型
 1. 語言獨立聲學模型與free syllable語言模型
 2. 語言獨立聲學模型與syllable bi-gram
 3. 語言相依聲學模型與free syllable
 4. 語言相依聲學模型與syllable bi-gram

實驗

- 在混合語言的**LID**實驗中,將辨識結果中相鄰為相同語言的語言標籤合併,形成一個語言區塊,用來計算正確率的單位量
- 而正確率的判斷評估方式分為兩種:
 - 1.語言標籤評估法
 - 2.語言時間資訊評估法

語言標籤評估法

- 分爲Hit、Deletion rate、Substitution rate、Insertion rate



圖五、語言標籤做爲評估單位範例

語言標籤評估法

- 這裡採用F-Measure做評估,F-Measure是依據精確率(Precision rate)及召回率(recall rate)做衡量標準

$$\frac{1}{F - m e a s u r e} = \frac{1}{2} \cdot \left(\frac{1}{p r e c i s i o n} + \frac{1}{r e c a l l} \right)$$

$$\Rightarrow F - m e a s u r e = \frac{2}{\frac{1}{p r e c i s i o n} + \frac{1}{r e c a l l}} = 2 \cdot \frac{p r e c i s i o n \cdot r e c a l l}{p r e c i s i o n + r e c a l l}$$

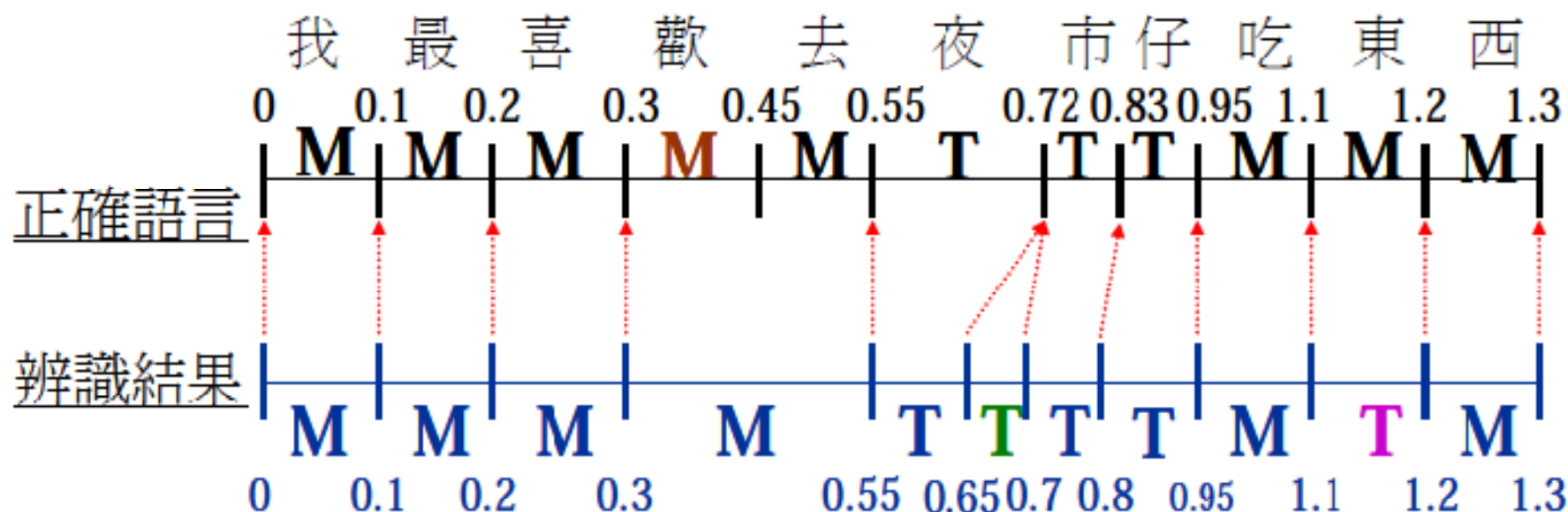
語言標籤評估法

表二、以語言標籤做為評估單位正確率

| | 以 Syllable 為語言計算單位 | | | 以單一語言片段 為語言計算單位 | | |
|-----------------------|-----------------------|--------|---------------|--------------------|--------|---------------|
| | P | R | F | P | R | F |
| LIAM+Free Syllable | 72.63% | 68.55% | 70.53% | 58.67% | 93.75% | 72.17% |
| LDAM+Free Syllable | 69.9% | 65.7% | 67.6% | 69.11% | 86.82% | 76.96% |
| LIAM+Syllable Bi-gram | 72.73% | 68.71% | 70.66% | 82.26% | 76.26% | 79.14% |
| LDAM+Syllable Bi-gram | 78.1% | 68.51% | <u>73.02%</u> | 82.35% | 79.0% | <u>80.67%</u> |

語言時間資訊評估法

- 首先對判別結果的語言時間邊界與正確答案語言時間邊界做Alignment

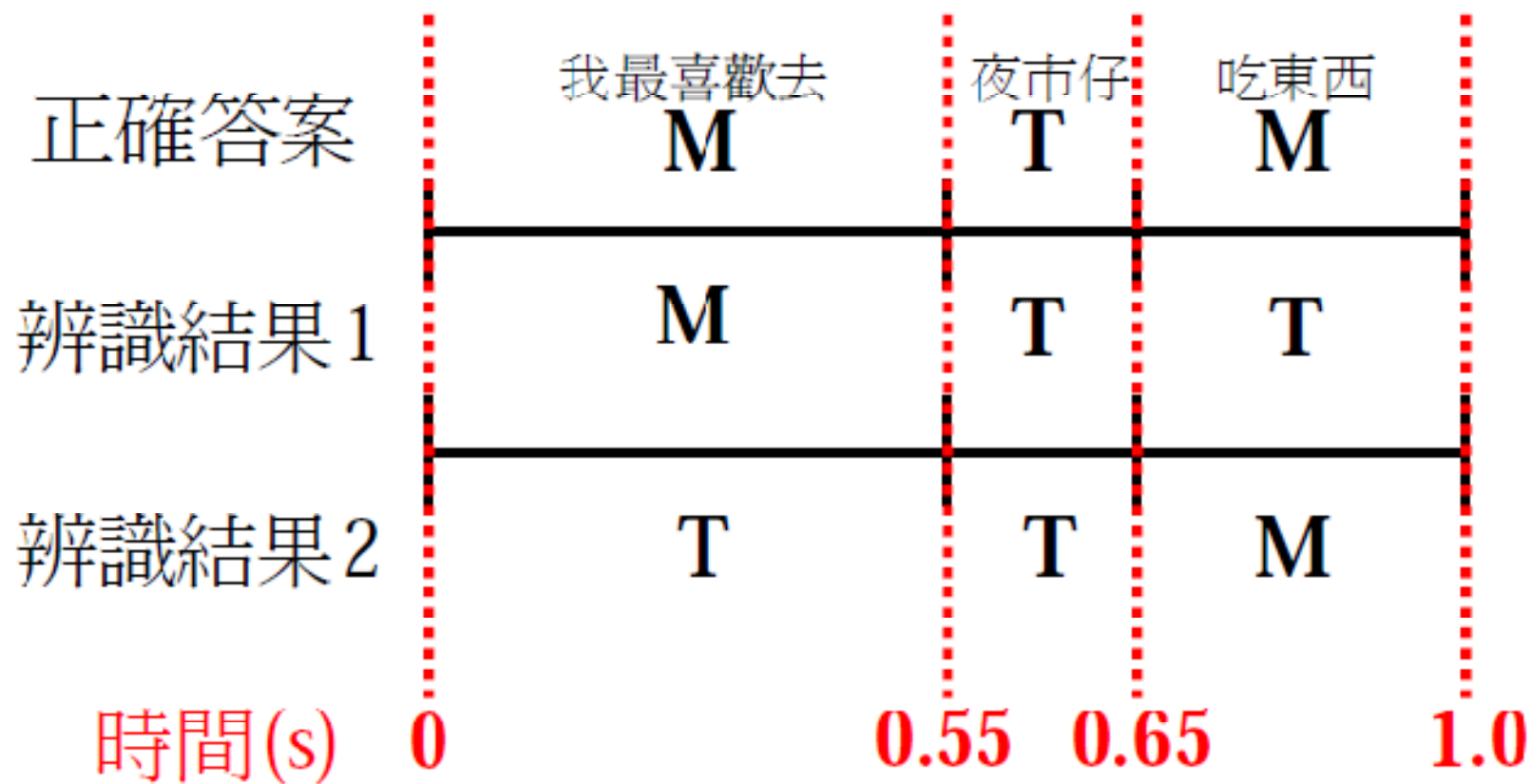


圖六、以語言時間資訊為評估單位

語言時間資訊評估法

- 以語言時間資訊為評估單位時,可以較清楚的了解判別錯誤的部分對整句語音的影響程度

語言時間資訊評估法



語言時間資訊評估法

表三、以語言時間資訊為評估單位正確率

| | 以 Syllable 為語言計算單位 | | | 以單一語言片段 為語言計算單位 | | |
|-----------------------|-----------------------|--------|--------------|--------------------|--------|---------------|
| | P | A | F | P | A | F |
| LIAM+Free Syllable | 75.3% | 68.68% | 71.84% | 60.2% | 52.78% | 56.25% |
| LDAM+Free Syllable | 77.1% | 68.87% | 72.75% | 68.8% | 58.05% | 62.97% |
| LIAM+Syllable Bi-gram | 84.1% | 70.76% | 76.8% | 79% | 60.57% | 68.57% |
| LDAM+Syllable Bi-gram | 84.6% | 71.76% | <u>77.3%</u> | 79.1% | 61.67% | <u>69.31%</u> |

實驗

- 在假設中語言轉換較常發生於word上,但如果要針對word做bi-gram則資料量需要太大,所以折衷計算部分較常出現的word與syllable間的bi-gram機率

實驗

表四、以語言標籤為評估單位正確率

| | 以 Syllable 為語言計算單位 | | | 以單一語言片段 為語言計算單位 | | |
|-----------------------|-----------------------|-------|--------------|--------------------|-------|--------------|
| | P | R | F | P | R | F |
| LDAM+Syllable Bi-gram | 78.1% | 68.5% | 73.0% | 82.4% | 79.0% | 80.7% |
| LDAM+Word&Syl Bi-gram | 86.4% | 63.3% | <u>73.0%</u> | 88.4% | 78.9% | <u>83.4%</u> |

表五、以語言時間資訊為評估單位正確率

| | 以 Syllable 為語言計算單位 | | | 以單一語言片段 為語言計算單位 | | |
|-----------------------|-----------------------|-------|--------------|--------------------|-------|--------------|
| | P | R | F | P | R | F |
| LDAM+Syllable Bi-gram | 78.8% | 84.4% | 81.5% | 69.2% | 79.0% | 73.8% |
| LDAM+Word&Syl Bi-gram | 78.2% | 86.7% | <u>82.2%</u> | 72.8% | 84.1% | <u>78.0%</u> |