語音關鍵詞擷取於自動影音新聞故事分段之應用

陳燦輝
國立台灣師範大學資訊工程研究所
g93470018@csie.ntntu.edu.tw

陳柏琳
國立台灣師範大學資訊工程研究所
berlin@csie.ntnu.edu.tw

王新民
中央研究院資訊科學研究所
whm@iis.sinica.edu.tw

摘要

隨著網際網路與多媒體技術的蓬勃發展，已經有許多的廣播電視公司開始將新聞影音檔放至網路上供使用者查詢與瀏覽，讓使用者能更直接清楚的了解新聞內容。目前的作法大多是以人工標記出某一則新聞文稿在影音檔中對應的時間點，讓使用者在閱讀新聞文稿的同時，可以點選對應的影音新聞，這樣的方式需要耗費大量的人力與時間，因此，只能選擇性地針對幾則重要新聞實施，而無法全面實現。有鑑於此，我們發展了一套自動找出新聞文稿在影音檔中對應段落的方法，我們先將網頁上每一則新聞文稿各自轉換成適當的關鍵詞，利用關鍵詞擷取(Keyword Spotting)技術，將每一個關鍵詞在一長段新聞語音中出現的位置依序標記出來，關鍵詞的起點位置，便是該則新聞的起點。我們以公視晚間新聞作為測試資料，進行實驗評估，實驗結果顯示，本論文提出的方法找出來的新聞片段，96%的起點偏移秒數小於 2 秒，精確率(Precision Rate)為 1，召回率(Recall Rate)則非常接近1，相當具有實用價值。

關鍵詞

影音新聞故事分段、關鍵詞擷取、語音辨識

1. 序論

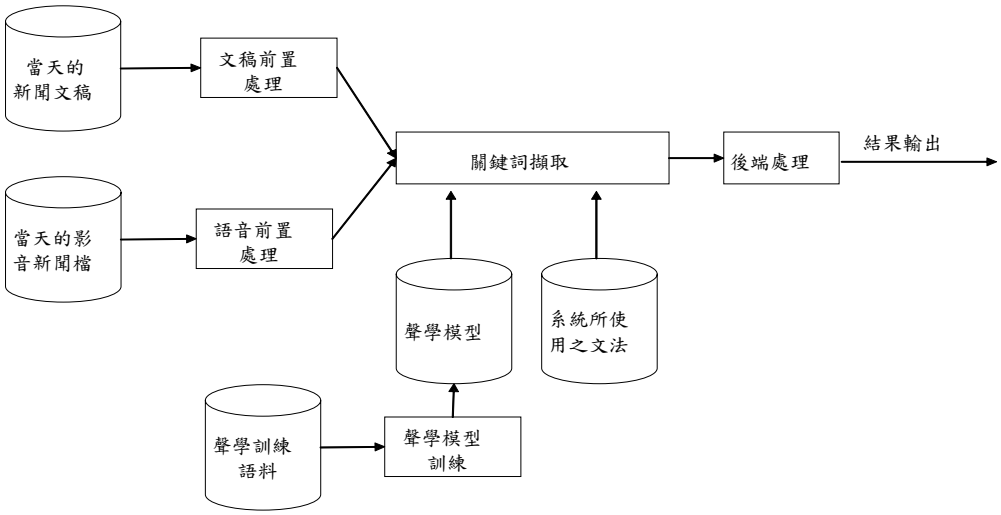
網際網路與多媒體技術的蓬勃發展，使得全球資訊網的內容也跟著越來越多元化，現在的網站，不再全部都只是文字，而是充斥著各式各樣的多媒體資訊，如圖片、影音及動畫檔等等，讓全球資訊網的內容變得生動活潑，不再是像以前一樣的刻板。對現在的人而言，縱橫網路似乎已經變成一項休閒娛樂，每個人上網的時間越來越長，也常利用網路查詢各項資訊。

新聞網站過去幾乎都只是將新聞的文字稿放至網路上供使用者瀏覽、查詢，對使用者來說，單單只是閱讀文字內容，並無法真正體會新聞的真實性，單純的文字敘述也不像影音檔，可以給使用者最直接的臨場感受。因此，如果能同時放置該則新聞文字稿對應的影音檔，這樣不但可以讓使用者直接看到新聞的現場實況，更了解新聞的實際情況，也必能大幅增加新聞網站的可看度，進而提昇對使用者的吸引力。目前已經有多家新聞媒體開始嘗試將每天的新聞影音檔放置在專屬網站供使用者觀看。只是到目前為止，大多是以人工標記出某一則新聞文稿在影音檔中對應的時間點，讓使用者在閱讀新聞文稿的同時，可以點選對應的影音新聞，這樣的方式需要耗費大量的人力與時間，因此，只能選擇性地針對幾則重要新聞實施，而無法全面實現。

有鑑於此，我們特別發展出一套完整的自動化標記系統，它可以自動找出新聞文稿在影音檔中的對應段落，也就是將新聞稿與影音新聞內容自動對齊。首先，將網頁上每一則新聞文稿各自轉換成適當的關鍵詞，然後，利用關鍵詞擷取(Keyword Spotting)技術，將每一個關鍵詞在一長段新聞語音中出現的位置依序標記出來，最後，關鍵詞的起點位置，便是該則新聞的起點。我們以公視晚間新聞作為測試資料，進行實驗評估，實驗結果顯示，本論文提出的方法找出來的新聞片段，96%的起點偏移秒數小於 2 秒，精確率(Precision Rate)為 1，召回率(Recall Rate)則非常接近1，相當具有實用價值。

另外，我們也將此方法與先前提出的非監督式(Unsupervised)影音新聞故事分段(Story Segmentation)[1]技術作一比較。這個方法先利用聲學特性改變偵測(Acoustic Change Detection)，將一長段音流(Audio Stream)自動切割成同質性較高的短音段，再利用音段分群(Audio Segment Clustering)技術，將上述的短音段依據聲音特性分群。假設最大的群對應到主播語音，而每一段主播語音都對應到一則新聞的起始。找出主播的語音段落後，便可以完成影音新聞故事分段。這個方法主要適用於只有影音新聞，並不存在新聞稿的情況。本論文提出的方法則是應用在已知新聞稿，想要在影音新聞中找出對應段落的情況，在進行影音新聞故事分段時有較多的資訊可供參考。這兩種方法都有其實用價值，所獲得的新聞故事分段結果可以作一比較。

本論文接下來的安排如下：第二章我們將說明本論文所提出的方法及概念，第三章我們簡單回顧先前的非監督式自動影音新聞故事分段方法，第四章描述我們的實驗環境及相關設定，並呈現我們的實驗結果及分析，第五章則為我們的結論及未來展望。



2. 新聞稿與影音新聞自動對齊系統

本研究的目標是在影音新聞中自動找出特定新聞稿的對應段落，也就是開發新聞稿與影音新聞自動對齊系統。相較於以前提出的非監督式影音新聞故事分段技術，這個方法也可以視為一種監督式影音新聞故事分段技術。如圖一所示，本論文所提出的方法其流程主要包括文稿前置處理、語音前置處理，聲學模型訓練(Acoustic Model Training)，關鍵詞擷取之文法(Grammar)設計，關鍵詞擷取(Keyword Spotting)，及後端處理等部份，各部份之核心技術將於以下各小節中詳細介紹說明。

2.1 文稿前置處理

本論文以公視新聞網[2]的公視晚間新聞為實驗標的。公視新聞網上頭的新聞文稿基本上是按日期排序，使用者輸入日期後，可以瀏覽當天的所有新聞。這些新聞文稿均為超文本標記語言(HTML)檔案，因此我們必須先將每一則新聞的超文本標記語言檔案中多餘的標籤(Tag)拿掉，只留下新聞文稿的本文及新聞標題。另外，在去除多餘的超文本標記語言標籤的同時，我們可以利用標籤上的資訊來對當天的新聞內容做適當的處理，例如將不是屬於公視晚間新聞的其它新聞文稿去除，最後計算出當天公視晚間新聞的實際故事則數，列出新聞的清單。

經過觀察，我們發現每一則新聞文稿的第一段幾乎都是主播的播報稿，第二段起，有時候則是外場記者與受訪者間的採訪稿。另外，有時候新聞稿會將主播、外場記者及受訪者的文稿全部混和，串成一大段。雖然根據經驗，關鍵詞擷取技術對於越長的關鍵詞有越高的鑑別力，但考慮系統執行的效率，以及新聞稿與新聞語音內容並不是百分之一百完全吻合，我們決定只取第一段最多前一百五十個中文字作為該則新聞的關鍵詞。

文稿前置處理的最後一步是找出上述關鍵詞的正確發音，我們採用中研院資訊所詞庫小組開發的斷詞程式(其詞典中共有 90,128 個詞，最短的詞為一字詞，最長的詞則為八字詞)替前一步驟取出的關鍵詞斷詞，便可以得到其中每一個字的正確發音。

2.2 語音前置處理

語音前置處理的目的是從語音信號中抽取適合語音辨識的聲學特徵參數。一般是先將輸入的語音經過在頻譜上的預強處理(Preemphasis，為一高通濾波器)，用以降低分佈在低頻的雜訊成分，同時強調分佈在高頻的語音成分。其次，聲學特徵參數抽取主要是在頻譜上作分析，求取出可以作為辨識依據的重要頻譜特徵向量。在本論文中，我們採用語音辨識常用的梅爾倒頻譜係數(Mel-frequency Cepstrum Coefficients, MFCC)特徵向量。在求取梅爾倒頻譜係數特徵向量時，我們先將語音資料切割成一連串部分重疊的音框(Frame)序列，而每一個音框得到由 13 維的梅爾倒頻譜係數特徵加上其一階與二階的時間軸導數(Time Derivatives)所形成共 39 維的特徵向量。其中 13 維的梅爾倒頻譜係數是由 18 個梅爾頻譜上濾波器組(Filter Banks)的輸出經餘弦轉換求得。為了降低錄音時的通道效應對語音辨識的影響，我們使用倒頻譜平均消去法(Cepstral Mean Subtraction，CMS)，嘗試在梅爾倒頻譜上消除通道效應的統計量。

2.3 聲學模型訓練

本論文中所使用的初始隱藏式馬可夫模型(Hidden Markov Model)是由四小時廣播新聞語音訓練而得。此廣播語料是經由收音機收集，為 1998 年 12 月至 1999

年 7 月之間台北地區數家廣播電台所播送之廣播新聞[3]。所有的錄音都經由人工切割成一則一則的新聞語音檔。內容只含新聞主播的語音，並不包括記者的現場採訪，至於主播的性別則是男女皆有。值得注意的是，部分檔案在收集時，因為錄音品質不佳而含有某些雜訊。

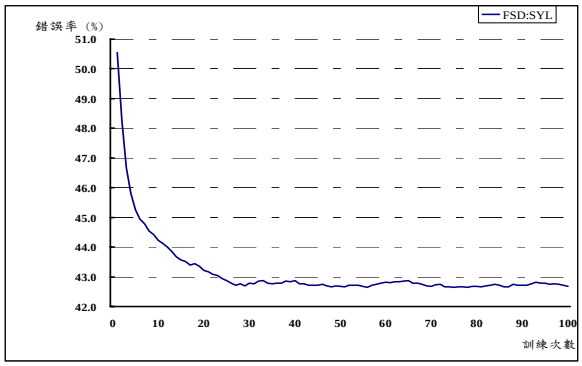
表一、聲學模型訓練語料統計資訊

性別	長度(分)	人數(人)
男	766.68	小於 66 人
女	766.68	小於 111 人

為了提高語音辨識的正確率，我們進一步採用 MATBN 電視新聞語料[4]，以前述的初始模型為基礎，進行聲學模型訓練。MATBN 為中研院資訊所中文資訊處理實驗室口語小組與公共電視台共花費三年的時間，合作錄製標記完成的新聞語音語料庫。錄製的對象是星期一至星期五每天晚間播出的公視晚間新聞，共收集標記了 198 小時的電視新聞語料。所有 MATBN 電視新聞語料都包含了正確的轉譯文字及有關說話者性別、類別(主播/記者/被採訪者)、背景聲音、及停頓、語助詞、呼吸、強調語氣、反覆及不適當的發音等資訊的詳細標記。所有標記及轉譯文字皆由專人使用 LDC&DGA 提供的 Transcriber[5]，按 LDC[6]訂定的標準格式完成。

考量到我們是將新聞稿的前面一百五十字作為關鍵詞，這些文字主要對應主播的語音，為了盡量避免在進行關鍵詞擷取時發生內部測試(Inside Test)的問題，影響實驗的客觀性，我們只選取外場記者(Field Reporter)的語音來訓練聲學模型。此外，為了使最後的聲學模型可以擁有較佳的語者獨立(Speaker Independent)及性別獨立(Gender Independent)特性，訓練語料也經過特別的篩選，聲學模型訓練語料統計資訊可參考表一。

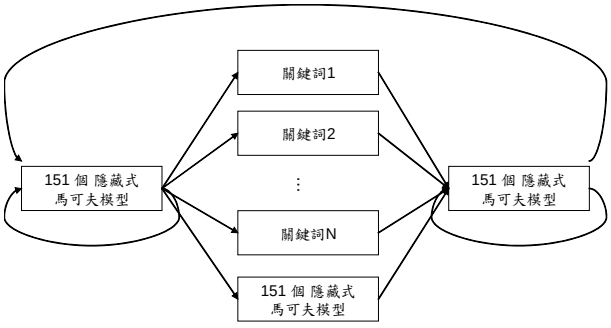
考慮中文語音結構，聲學模型由 22 個聲母(Initial)、38 個韻母(Final) 及一個靜音(Silence)模型組成，其中聲母模型根據右邊緊鄰的韻母種類可再進一步細分成 112 個模型，因此，最後我們共訓練出 151 個隱藏式馬可夫模型來作為做關鍵詞擷取時所使用之聲學模型[3]。在隱藏式馬可夫模型中，每個狀態(State)會依據其對應到的訓練語料多寡，用 2 到 128 個高斯統計分佈來表示。而訓練的方式，則是採用目前一般最常使用的最大相似度(Maximum Likelihood)訓練，而我們訓練的次數從 1 次到 100 次，在未有任何語言模型輔助的情況下，自由音節辨識(Free Syllable Decoding)獲得的音節錯誤率如圖二所示。我們最後決定採用訓練 100 次後的隱藏式馬可夫模型作為關鍵詞擷取的聲學模型。



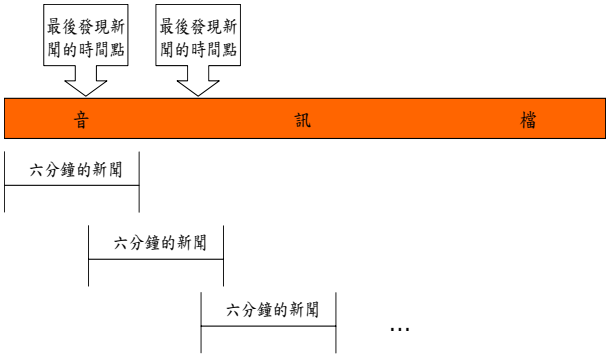
圖二、最大相似度訓練的自由音節辨識之音節錯誤率曲線

2.4 搜尋文法設計

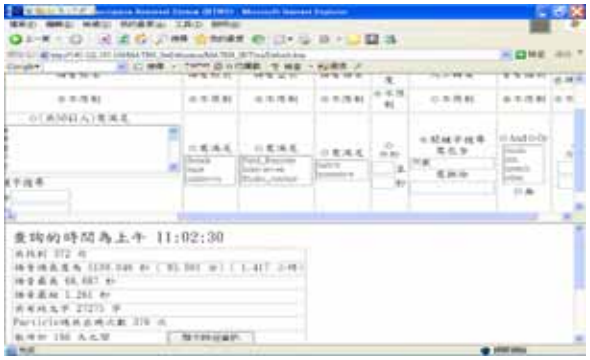
我們採用 HTK[7]工具庫中的 HVite 函式來進行關鍵詞擷取。我們也利用 HTK 工具庫裡面的另一項工具 HParse 設計一套文法，來限制關鍵詞擷取程式在進行隱藏式馬可夫模型狀態層次(State-level)維特比動態規劃搜尋(Viterbi Dynamic Programming Search)時的合法路徑，進而達到更好的效果。我們使用的文法如圖三所示。我們先利用文稿前置處理的輸出，將每一則新聞稿中文字部分，根據發音，將對應的聲母/韻母模型串起來，至於出現標點符號的地方，則插上一個可選擇性出現(Optional)的靜音(Silence)模型。關鍵詞擷取需要一組填充模型(Filler Model)來對應非屬關鍵詞的語音段落，我們的方法是直接採用 151 個隱藏式馬可夫模型作為填充模型，並利用每次模型轉移(Model Transition)時加上懲罰分數(Penalty)的方式，避免關鍵詞出現的地方被那 151 個隱藏式馬可夫模型串成的音節(Syllable)串列取代，產生錯誤拒絕(False Rejection)。



圖三、關鍵詞擷取所用之文法



圖四、基於關鍵詞擷取之影音新聞故事分段示意圖



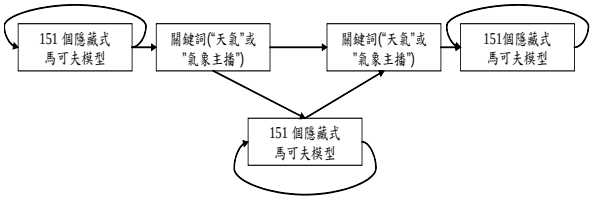
圖五、MATBN 電視新聞語料資訊檢索系統

2.5 關鍵詞擷取

在做關鍵詞擷取時，我們並沒有加上語言模型(Language Model)的資訊。透過一些簡單的實驗觀察，我們發現隱藏式馬可夫模型轉移的懲罰分數(Penalty)在設定為 50 的時候可以得到不錯的結果。由於 Hvite 函式並無法一次處理長達約一個小時的新聞，因此，如圖四所示，我們設計的系統處理流程是一次處理 6 分鐘長度的音訊片段，完成關鍵詞擷取之後，找出在這 6 分鐘片段中所有出現的關鍵詞，從最後一個關鍵詞開始處再取出 6 分鐘片段進行關鍵詞擷取，如果沒有找到關鍵詞的話，則往後取出 6 分鐘片段進行關鍵詞擷取。

2.6 後端處理

後端處理的主要工作是找出最後一則新聞的結束點及處理因為新聞文稿錯誤造成的新聞故事分段點時間偏移。由於我們是利用每則新聞最多前 150 個中文字當成是關鍵詞，因此關鍵詞擷取的結果，只能知道每則新聞的開始時間，卻無法得知新聞的結束點。不過，由於新聞播報的特性是下一則新聞的開始點便是前一則新聞的結束時間，所以對大多數的新聞故事來說，問題不大，但這個方法無法找出最後一則新聞的結束時間。由於國內一般的新聞節目都會在最後安排氣象預報，如果可以將氣象預報偵測出來，將氣象預報的開始視為最後一則新聞的結束，應是可行。我們利用國立台灣師範大學資訊工程研究所郭人瑋先生開發的 MATBN 語料資訊查詢系統[8](如圖五所示)對 198 小時新聞的轉譯文字進行觀察，發現「天氣」及「氣象主播」兩個詞是氣象預報開始的時候常見的兩個詞彙。我們利用關鍵詞擷取，找出所有新聞的起點之後，再選用「天氣」及「氣象主播」兩個關鍵詞，在最後一則新聞的起點至當天新聞最後的結束點間的音段，進行關鍵詞擷取。圖六說明這一次關鍵詞擷取使用的文法，我們強迫這兩個關鍵詞在這個音段中只能出現兩次，而每一次出現關鍵詞的時間點，可以任意選擇要出現「天氣」或「氣象主播」，最後，我們以關鍵詞



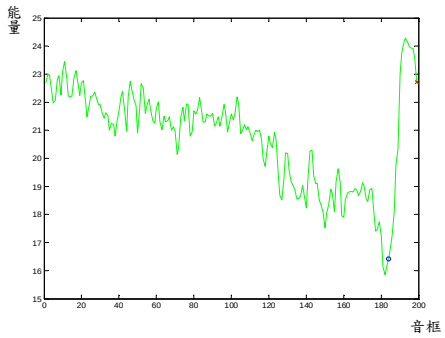
圖六、尋找最後一則新聞結束點的文法

第一次出現的時間點視為最後一則新聞的結束時間點。

另外，在實驗過程中，我們發現有時候會因為新聞文稿的輸入錯誤，或是主播並非完全按稿播報，造成以關鍵詞擷取技術找到的新聞開始時間，其起點偏移秒數偏高，或是讓使用者點選影音新聞時，感覺很突兀。例如，一則影音新聞的開頭其實是「陳水扁總統」，但是新聞稿的開頭卻是「陳總統」三個字，關鍵詞擷取的結果，將文字的「陳總統」對應到「扁總統」的語音段落，使用者點選這一則新聞的影音檔後，將聽到由「扁總統」三個字開始的新聞。雖然對起點偏移秒數的計算來說，這些誤差僅是微不足道的零點幾秒，但對使用者來說，這一點小小的差距就可能造成其感官上的極不滿足。針對此點，我們提出了一個解決辦法。根據觀察，大部份新聞在結束之後，往往伴隨一小段靜音，然後才是下一則新聞的開始。對前面的例子而言，我們若將關鍵詞擷取找到的新聞起點往前兩秒鐘，以音框長度 20ms，重疊 10ms 方式，描畫出近 200 個音框的時域能量曲線(Energy Contour)，會得到圖七的結果。我們發現，真正的新聞起點非常靠近能量波谷處。所以，我們可以從關鍵詞擷取找到的新聞起點，往前找出能量曲線中的第一個波谷對應的時間點，作為修正後的新聞起點。

3. 非監督式影音新聞故事分段技術

本論文提出的影音新聞故事分段方法，將與我們以前提出的非監督式影音新聞故事分段方法作效能評比。這個非監督式影音新聞故事分段法先利用聲學特性改



圖七、關鍵詞擷取找到的新聞起點(紅點)往前 2 秒間的時域能量曲線(藍點為正確起點)

變偵測(Acoustic Change Detection)，將一長段音流(Audio Stream)自動切割成同質性較高的短音段，再利用音段分群(Audio Segment Clustering)技術，將上述的短音段依據聲音特性分群。假設最大的群對應到主播語音，而每一段主播語音都對應到一則新聞的起始。找出主播的語音段落後，便可以完成影音新聞故事分段[1]。以下是這個方法各個步驟的介紹：

3.1 基於貝氏資訊基準之模型選擇

貝氏資訊基準(Bayesian Information Criterion, BIC)可以用來從一組候選模型(Candidate Model)中選擇一個最適合的模型，來描述一組已知的資料集(Data Set)。令 $X=\{x_1,x_2,\cdots,x_N\}$ 為一組我們欲描述的資料集，$M=\{M_1,M_2,\cdots,M_k\}$ 為候選模型組，模型 M_i 的 BIC 值可以定義成：

$$BIC(M_i)=\log L(X,M_i)-\lambda\frac{1}{2}\#(M_i)\times\log(N) \qquad (1)$$

其中 $L(X,M_i)$ 代表資料 X 在模型 M_i 下的最大相似度，$\#(M_i)$ 為模型 M_i 的參數量，N 為資料集的個數，λ 則是補償加權(Penalty Weight)，其值通常都設為 1。獲得最大 BIC 值的模型便是表示資料 X 的最佳模型，

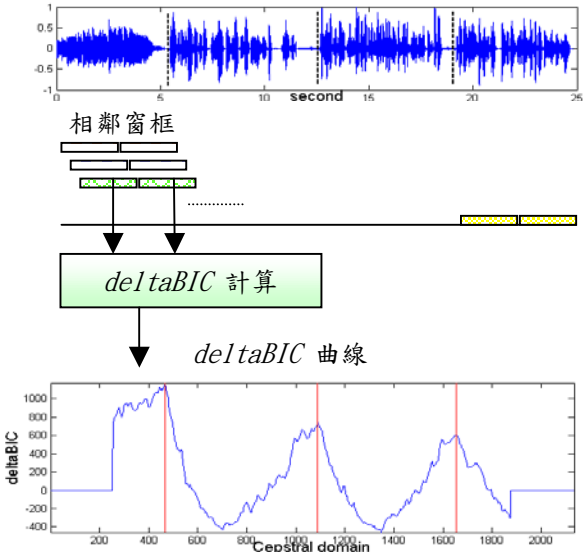
$$M_x=\arg\max_{M_i}\left[\log L(X,M_i)-\lambda\frac{1}{2}\#(M_i)\times\log(N)\right] \qquad (2)$$

3.2 基於貝氏資訊基準之聲學特性改變偵測

經過語音前置處理，一段音流(Audio Stream)將被表示成一個特徵向量序列，如圖八所示，藉由計算相鄰兩個窗框(Window)的 *deltaBIC* ，可以得出一條 *deltaBIC* 曲線。*deltaBIC* 定義如下：

$$deltaBIC=BIC(M_1)-BIC(M_0) \qquad (3)$$

(3)式比較兩個模型：M_1 代表兩個窗框的特徵向量必須分別用一個多變量高斯分佈(Multivariate Gaussian)來描述，而 M_0 則表示兩個窗框的特徵向量可以共用一個多



圖八、基於貝氏資訊基準之聲學特性改變偵測法

變量高斯分佈。如果 $deltaBIC<0$，就表示兩個窗框聲學特性近似，相反地，如果 $deltaBIC>0$，就表示窗框聲學特性不同。所以 *deltaBIC* 曲線上的峰點便對應到音訊的特性轉換處。

3.3 基於貝氏資訊基準之分群法

假設 $S=\{s_1,s_2,\cdots,s_L\}$ 是我們欲分群的音段，每一個音段分別表示成一個特徵向量序列，如果給定兩個音段 s_i 及 s_j，我們比較兩個模型：M_1 代表兩個音段的特徵向量必須分別用一個多變量高斯分佈描述，而 M_0 則表示兩個音段的特徵向量可以共用一個多變量高斯分佈。採用式(3)，如果 $deltaBIC<0$ 就表示兩個音段可以合併到同一群。先假設每一個音段自成一群，由下而上(Bottom-Up)合併，直到所有的群都不能合併為止。

4. 實驗結果

我們從有正確標記資訊的公視晚間新聞語料中任意抽出五天的新聞進行實驗評估。

4.1 二種自動新聞故事分段方法之比較

第一個實驗是要比較基於貝氏資訊基準的非監督式新聞故事分段及本論文所提出的自動新聞故事分段的效果。在這邊，我們參考文獻[1]的評估方法，表二是使用基於貝氏資訊基準的非監督式新聞故事分段的結果，而表三則是使用本論文所提出方法的結果。根據公共電視新聞網站上的資料，測試語料採用的五天晚間新聞共有 84 則報導。這 84 則新聞中，有 2 則與前一則新聞間並沒有間隔其他記者或是受訪者的語音，

表二、使用基於貝氏資訊基準的非監督式影音新聞故事分段的結果

			第一天	第二天	第三天	第四天	第五天
實際新聞則數			16	18	18	15	17
偵測到的新聞則數			16	18	17	15	17
錯誤拒絕則數			0	0	1	0	1
錯誤接受則數			0	0	0	0	0
起點偏移秒數	小於 2 秒	之則數	15	18	14	13	14
	2 到 3 秒		0	0	0	0	0
	大於 3 秒		1	0	3	2	2
起點總偏移秒數			12.29	6.6	20.67	37.66	30.48

表三、使用本論文所提出的影音新聞故事分段方法的結果

			第一天	第二天	第三天	第四天	第五天
網路上的文稿新聞則數			16	18	18	15	17
偵測到的新聞則數			16	18	17	15	17
錯誤拒絕則數			0	0	1	0	0
錯誤接受則數			0	0	0	0	0
起點偏移秒數	小於 2 秒	之則數	15	18	15	15	17
	2 到 3 秒		0	0	1	0	0
	大於 3 秒		1	0	1	0	0
起點總偏移秒數			7.28	4.86	11.60	5.34	5.46

會被基於貝氏資訊基準的非監督式新聞故事分段法視為與前一則新聞是同一則，不會被偵測出來，所以此方法共分出 82 則，其召回率(Recall Rate)跟精確率(Precision Rate)分別為 0.976(82/84)和 1.0(82/82)。表二顯示，在被正確切出的 82 則新聞中，共有 74 則(90%)的起點偏移秒數在 2 秒內，0 則新聞在 2 至 3 秒之間，而剩下的 8 則新聞，其偏移秒數大於 4 秒，甚至有 3 則新聞起點偏移秒數大於 10 秒。

表三顯示，使用本論文提出的自動新聞故事分段方法，84 則新聞中，有 83 則會被正確的偵測出來，因此其召回率為 0.988(83/84)，而精確率為 1.0(83/83)，而被正確偵測的 83 則新聞中，共有 80 則(96%)新聞的時間偏移秒數在 2 秒內，1 則在 2 至 3 秒之間，而剩下的 2 則新聞，其起點偏移秒數分別為 3.9，4.5 秒。比較表二及表三，本論文提出的方法在起點時間偏移秒數方面明顯優於基於貝氏資訊基準的非監督式新聞故事分段，兩個方法的精確率都是 1，在召回率方面，本論文提出的方法也略優於基於貝氏資訊基準的非監督式新聞故事分段。

4.2 偵測最後一則新聞結束時間之方法評估

實驗二主要是評估最後一則新聞結束時間偵測的效果。評估的方式則是看當天的新聞正確標記中，最後一則新聞的結束時間與自動偵測的結果間的偏移秒數。由表四可以發現，目前採用的簡單方法輸出的結束點與正確的結束點間偏移秒數平均為 140 幾秒，如果不考慮誤差將近 600 秒的那一天，其餘四天的誤差還算可以接受。由於播放期間使用者可以自行按下結束按鈕，一般而言，新聞結束時間的精確與否，不像開始時間那麼關鍵。雖然這個部分的誤差很大，還是

表四、最後一則新聞結束點偏移秒數

	結束點偏移秒數
第一天	16.00
第二天	93.32
第三天	9.67
第四天	589.07
第五天	6.67
總結束點偏移秒數	714.73

表五、考慮時域能量曲線來降低起點偏移秒數誤差的結果

	原本起點總偏移秒數	考慮時域能量曲線後之起點總偏移秒數
第一天	7.28	5.38
第二天	4.86	2.88
第三天	11.60	9.56
第四天	5.34	2.96
第五天	5.46	5.20
起點總偏移秒數	34.36	25.98

有一定的用處，不過，這個部分顯然有很大的改進空間。

4.3 以時域能量曲線修正新聞偏移誤差之探討

當新聞稿有一點小錯誤，特別是當錯誤是發生在新聞稿的開始處，往往會造成新聞分段點的偏移誤差，降低使用者觀賞影音新聞時的滿意度。實驗三的目的是測試時域能量曲線是不是可以作為修正新聞起點的依據，如表五所示，參考時域能量曲線修正新聞偏移誤差之後，測試語料每一天的新聞起點總偏移秒數都下降，效果相當顯著。

5. 結論及未來展望

根據實驗結果可得知本論文提出的方法，不論是召回率或是精確率，甚至是新聞開始時間的偏移秒數，都有令人滿意的結果。相較於過去的非監督式影音新聞故事分段法，本論文提出的方法由於多使用了新聞文稿的資訊，可以大幅降低新聞分段點的誤差偏移秒數。未來的研究方向主要包括：1. 改良最後一則新聞結束點偵測，目前本論文所提出的解決辦法相當粗略，未來顯然有很大的改進空間，比如說與非監督式影音新聞故事分段結合，或是設計一個更好的關鍵詞擷取文法。2. 不同的新聞媒體在新聞格式及編排上可能與公視新聞不同，我們需要進行更廣泛的實驗評估。3. 仿照新聞媒體網站，完成實體展示系統，並希望與新聞媒體合作，直接將系統與新聞網站結合。

6. 致謝

這篇論文能夠完成要感謝中研院資訊所鄭士賢先生及台師大資工所語音實驗室郭人瑋先生的熱心幫忙和提供資料。

7. 參考資料

[1] H.M. Wang, S.S. Cheng, and Y.C. Chen, “The SoVideo Mandarin Chinese Broadcast News Retrieval System,” International Journal of Speech Technology, Vol. 7, No. 2-3. pp. 189-202, 2004.

[2] 公視新聞網網址：http://www.pts.org.tw/php/news/new_main.php/

[3] B. Chen, H.M. Wang, and L.S. Lee, “Discriminating Capabilities of Syllable-Based Features and Approaches of Utilizing Them for Voice Retrieval of Speech Information in Mandarin Chinese,” IEEE Trans. on Speech and Audio Processing, Vol. 10, No. 5, pp. 303-314, 2002

[4] H.M. Wang, B. Chen, J.W. Kuo and S.S. Cheng, “MATBN: A Mandarin Chinese Broadcast News Corpus,” International Journal of Computational Linguistic and Chinese Language Processing, Vol. 10, No. 2, pp. 219-236, 2005.

[5] C. Barras, E. Geoffrois, Z.B. Wu, and M. Liberman, “Transcriber : Development and Use of a Tool for Assisting Speech Corpora Production,” Speech Communication, Vol. 33, pp. 5-22, 2001.

[6] Linguistic Data Consortium : http://www.ldc.upenn.edu.

[7] S. Young, G. Evermann, D. Kershaw, G. Moore, Julian Odell, D. Ollason , D. Povey, V. Valtchev and P.C. Wooland. The HTK Book. Version 3.2, 2002. http://htk.eng.cam.ac.uk/

[8] 國立台灣師範大學資訊工程研究所，語音訊號處理實驗室. http://speech.csie.ntnu.edu.tw/