

行政院國家科學委員會專題研究計畫 成果報告

中文自發性語音語料庫之建立(3/3)

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-009-021-

執行期間：92年08月01日至93年07月31日

執行單位：國立交通大學電信工程學系

計畫主持人：陳信宏

共同主持人：王小川，鄭秋豫，吳宗憲，王新民，王駿發，李琳山

計畫參與人員：張森嘉，陳榮貴

報告類型：完整報告

處理方式：本計畫涉及專利或其他智慧財產權，1年後可公開查詢

中 華 民 國 93 年 11 月 9 日

中文自發性語音語料庫之建立(3/3)

Spontaneous Mandarin Speech: Corpus and Processing

期末報告

計畫編號：NSC-92-2213-E-009-021

執行期限：92年8月1日至93年7月31日

全程計畫：90年8月1日至93年7月31日

主持人：陳信宏 國立交通大學電信工程學系
schen@mail.nctu.edu.tw
共同主持人：李琳山 國立台灣大學電機工程學系
王小川 國立清華大學電機工程學系
王駿發 國立成功大學電機工程學系
吳宗憲 國立成功大學資訊工程學系
鄭秋豫 中央研究院語言學研究所
王新民 中央研究院資訊科學研究所
張森嘉 工研院前瞻研究中心
陳榮貴 中華電信研究所

目 錄

中文摘要	3
Abstract	4
1. 緣由與目的	5
2. 計畫內容	7
2.1 語料收錄	7
2.2 語料文字標註處理	12
3. 結果與討論	19
3.1 語料收錄及標註結果	19
4. 結論	22
致謝	23
參考文獻	24
附件	26
附件一：電視新聞錄音資料處理技術報告	27
1. 原始資料	28
1.1 來源	28
1.2 節目	28
1.3 錄音/錄影配備	28
2. 資料保存	28
2.1 聲音資料	29
2.2 影音資料	35
3. 標註	40
3.1 基本架構	40
3.2 語音資料標誌	43
4. 實際操作問與答	55
附件二：電視節目錄音資料及對話語音資料處理技術報告	61
1. 錄音資料來源	62
2. 使用軟體	62
2.1 CoolEdit 2000	62
2.1.1 CoolEdit 2000 功能簡介	62
2.1.2 CoolEdit 基本使用說明	63
2.2 Transcriber	65
2.2.1 Transcriber XML 架構介紹	65
2.2.2 XML 標籤介紹	66
2.2.3 Type 詳細介紹	67
2.2.4 Transcriber 使用說明	69
2.3 UltraEdit	75
2.3.1 UltraEdit 功能簡介:	75
3. 語音資料標示成果	75
3.1 IC 電台標註狀況	75
3.2 中央社標註狀況	77
3.2.1 對話語料 2 中央社製播	77
3.2.2 對話語料 3 中央社製播	85

中文摘要

關鍵詞：自發性華語語音資料庫、新聞廣播語音、對話語音、文字標註、自發性語音辨認

本計畫共收集了四種華語自發性語音資料，包括新聞廣播語音、電台廣播語音、日常對話語音、及仿人機對話語音，並對其中的新聞語音、電台廣播語音、及日常對話語音三個語料進行人工文字標註處理，完成了 198 小時的廣播新聞語音處理；8 小時 36 分 12 秒的電台廣播語音處理，含 70616 字；及 3 小時 23 分 30 秒的日常對話語音處理，含 50566 字。本報告說明各類語料的內容、文字標註使用的軟體工具及處理做法、以及主要的成果。本計畫產出的語料庫將提供國內外學術界，作為自發性語音辨認研究之用。

Abstract

Keywords: Spontaneous Mandarin speech database, Broadcast news speech, Dialog speech, Transcription, Spontaneous-speech recognition

In this project, we collect four types of spontaneous Mandarin speech. They include broadcast news, radio program, human dialog speech, and simulated man-machine dialog speech. Processings to transcribe and segment speeches of the first three types have been done. Three databases are hence constructed. One is a 198-hour broadcast news database. Another is an 8.6-hour radio program speech database. The other is a 3.4-hour dialog speech database. They will be used in spontaneous Mandarin speech recognition.

第一章 緣由與目的

語音辨認技術的研究發展已經有數十年的歷史，過去幾年，藉由大量的語料訓練出精細的語音模型(Acoustic Model)及語言模型(Language Model)，大字彙(Large-Vocabulary) 朗讀式連續語音 (Continuous Read Speech) 辨認研究已有顯著的進步。國際上，學術研究方面最具代表性的包括美國的麻省理工學院(MIT)、卡耐基美隆大學 (CMU)、英國的劍橋大學及其他歐陸的一些學校，研究機構與業界方面則包括美國 IBM、Microsoft、BBN、日本 ATR、NTT、香港 Infotalk、荷蘭 Philips 等。

中文語音辨認研究方面，國內學術界以台、清、交、成大、中央及中研院等研究團隊為主，近年也有一些新興學校或年輕學者開始投入，研究機構包括工研院電通所、電信研究所、資策會等，許多業界也開始積極投入，如宏碁、蒙恬、聲碩、台達電、語博及英泰等，而中國大陸亦有以北京清華大學及中科院為主的許多研究團隊，值得注意的是一些跨國公司如 IBM、APPLE、Microsoft、Philips 等近年在中國大陸市場的刺激下，亦紛紛投入中文語音處理研究；以北京的 Microsoft 研究中心為例，他們收集了 3000 人（遍及大陸主要區域及台灣），每人 200 句的中文語音，以發展其辨認系統，。

目前，國際上大字彙語音辨認研究的重心已移轉到真實世界存在的自發性語音 (Real World Speech, or Spontaneous Speech)，例如廣播電視新聞 演講等。1995 年，DARPA 公告以廣播新聞 (Broadcast News) 語料為主的一套新的語音辨認評比標準。1996 年，NHK Science & Technical Research Laboratories 則與日本境內的研究機構及大學合作展開電視新聞語音辨認的研究計畫。日本更在 1999 年開始進行一個為期 5 年的國家型計畫，擬收集 1000 小時的自發性語音，將其中的 100 小時語音 transcript，並標示 word boundary、詞類及韻律標記（使用 J-ToBI），剩下的 900 小時語音則只 transcript，並標示 word boundary 及詞類。近年來，歐陸一些國家甚至已朝向多語言(Multilingual) 新聞語音處理研究。資訊檢索(Information Retrieval)界的年度盛會 TREC (Text REtrieval Conference) 從 1996 年起已將語音文件檢索(Spoken Document Retrieval)納入評比的項目，檢索標的即是新聞語音資料。The Topic Detection and Tracking Evaluation Project (TDT)

自 1997 年第一屆起即將報紙新聞與廣播語音新聞同時納入主題偵測與追蹤 (Topic Detection and Tracking) 技術評比，自 1998 年第二屆起更將中文一併納入，進行跨語言實驗評比。

國際上各相關研究機構均已積極投入這類極具前瞻性之廣播新聞語音辨認與資訊檢索相關研究，而這類研究之所以能如此蓬勃發展，負責大量語料收集與整理工作的 LDC (Linguistic Data Consortium) 實厥功至偉，否則動則上百小時的語料絕非一般研究團隊有能力自行建立之實驗環境。國內雖缺乏類似 LDC 這樣一個負責語料收集的專責機構，由國科會贊助之 MAT 三年期語料收集計畫在清大王小川教授主持下，由台大、成大、中興、中央、淡江、交大、中研院、電信研究的研究團隊積極配合，已克服重重困難完成電話環境下之語料收集階段任務，而 MAT 語料也確實在國內各研究團隊發揮其應有的功能，MAT 計畫可謂創下非常成功的先例。

因此，本計畫的目的是延續上一個錄音計畫的精神，由各研究團隊合作，進一步收集華語自發性語音，包括新聞語音、電台廣播語音、日常對話語音、及仿人機對話語音，經以人工做發音文字標註，建立實驗語料庫，供國內語音辨認與資訊檢索等研究團隊使用，俾使國內語音辨認及資訊檢索研究早日臻於世界一流水準，未來也可提供國外學術界使用，建立國際交流，提升華語研究水準。

第二章 計畫內容

本計畫共收錄了四種華語自發性語音語料：新聞語音、電台廣播語音、日常對話語音、及仿人機對話語音，並對其中的三個主要語料—新聞語音、電台廣播語音及日常對話語音進行人工文字標註處理，以建立實驗語料庫，供國內語音辨認與資訊檢索相關研究團隊使用。以下對四種語料收錄及三個語料庫的處理分別加以說明：

2.1 語料收錄

我們共收錄了新聞語音、電台廣播語音、日常對話語音、及仿人機對話語音四種華語自發性語音語料，語料來源為公共電視台、中央社及新竹 IC 電台，其中由公共電視台錄製了所有的新聞語音，由中央社錄製了部分的電台廣播語音、所有的日常對話語音及仿人機對話語音，由新竹 IC 電台錄製的為部分的電台廣播語音。對錄製五種語料的相關資訊分別說明如下：

(1) 新聞語音之錄製

所有的新聞語音都在公視錄製的，在第一年度計畫執行之初的數個月先進行各項準備工作，包括聯繫電視及廣播公司洽談授權、準備標註軟體、決定標註方式等。經與公共電視洽談後，公視同意授權我們使用其新聞節目，並建議我們採用『公視新聞深度報導』節目及願意協助我們錄音（影），錄音工作自 90 年 11 月 7 日起正式展開。『公視新聞深度報導』於每週一至五晚間 21:00-22:00 播出一個小時，自 91 年 7 月起，變更節目名稱為『公視晚間新聞』，自 91 年 9 月起，播出時間改為晚間 21:00-21:45，播出 45 分鐘，另於 21:45-22:00 播出 15 分鐘的『公視手語新聞』，92 年 1 月 31 日起，『公視晚間新聞』移至 19:00-20:00 播出，21:00-21:45 則播出『公視全球現場』，21:45-22:00 仍播出『公視手語新聞』。自 90 年 11 月 7 日起至 92 年 2 月底止，錄音時間固定為 21:00-22:00，92 年 3 月起，錄音時間則包括 19:00-20:00 及 21:00-22:00 兩個時段。本計畫錄音工作進行至 92 年 6 月底結束，共收錄約 300 個小時的新聞節目，主要內容為國內新聞，也有一小部分為國際新聞。語料收集及語料保存方式簡單說明如下：

A. 語料收集

- (a) 錄音採 TASCAM DA-40 DAT 錄音座，經由主控台在新聞播放時利用 AES/EBU 平衡式類比輸入同步錄音。
- (b) 錄影採 SONY SLV-ED88 錄放影機，利用一般 RCA 接頭同步錄影，錄影帶採用 TDK HS-160 型號。
- (c) 錄音/錄影格式
 - ◆ DAT tape: 格式：44.1kHz、16bit、stereo
 - ◆ VHS tape: stereo

B. 語料保存

(a) 聲音資料

- ◆ 公視取回的 DAT(數位錄音帶)，經 USB 介面直接將錄音帶內的數位信號讀進 PC 內轉為格式為 44.1kHz、16bit、stereo 的聲音檔 (windows PCM、.wav)，並燒錄於光碟中以便保存。
- ◆ 標註使用的聲音檔，因考量檔案傳輸及讀取速度的問題，將原始的檔案，利用聲音編輯軟體 — CoolEdit 2000 將已轉為 windows PCM 的聲音檔進行格式轉換。轉換為 16kHz、16 bit、mono 後，為便利日後管理及利用，每週的公視新聞深度報導儲存於同一光碟中保存。

(b) 影音資料

公視取回的 VHS 錄影帶，經由 UPMOST 301BTR 類比影像擷取卡，擷取 avi 格式的影像，並由影像編輯軟體 — 會聲會影(友立出品)即時壓縮成 MPEG1 格式保存。

(2) 電台廣播語音之錄製

電台廣播語音有兩個來源：中央社及新竹 IC 電台，其中中央社語音為訪談性節目，直接由中央社內部錄音提供；新竹 IC 電台語音為交談節目及評論性新聞，由收音機錄下無線廣播語音，分別說明如下：

A. 中央社語音

從 2002 年底開始到 2003 前五月，陸續從中央社取得電台訪談的語料(屬於自發性 Spontaneous Speech)，有九片光碟(共 77 個檔案，合約九小時又三十分

鐘，內容以第一片為範例見附錄一)，實付中央社八萬元台幣。在我們的要求下，每個檔案錄音格式為標準的 Windows wav (Linear PCM¹⁶)，也附有描述每段訪談的主題和對話者的性別的文字檔。訪談內容以電台廣播方式呈現，內容大多為男、女主播採訪著名專業人士，剖析其相關領域的專業知識，有的檔案有獨立的主題，有些則是具連貫性主題的系列訪談。大部分的訪談錄音中，說話速度適中，語調清晰，但是部份檔案音量偏小，主播或受訪者聲音含混，也有穿插笑聲、彼此搶話等現象。

由於主播及被訪者的交談屬於自發性語音（並非照著稿子所唸），因此常常出現各式各樣的語氣詞、停頓、口語不清、喘氣、搶話、笑聲、聲量忽大忽小、發音不清等現象，例如連續說"對對對對對"來贊同對方所講的話，發出"嗯"、"哦"、"那"、"哇"等語氣詞；在多人訪談時，時常會有搶話的現象；而節目開始前通常會先來段音樂緩和氣氛或是訪談到一半大家想休息一下時，也會播放音樂。

B. 新竹 IC 電台語音

經過和新竹 IC 電台洽談後，我們獲得他們的許可錄製語料，經處理去除牽涉個人隱私語料後，我們可以使用處理後之語料。因此我們開始由廣播直接錄製訪問性語料，經先處理一小部份語料進行語音辨認後，確認如此錄製之語料可以使用，之後我們共錄製了五個節目，內容類別說明如下：

(a) 歡喜一家親、愛上新竹、心靈故鄉

皆是訪談性節目，主持人及來賓對於某一主題做深入討論。

(b) IC 晚報

前半播報新聞，後半有主播和記者...等對於一些新聞的討論。

(c) IC 三朵花

由兩三個人輪流出來談一些生活常識。

(3) 日常對話語音之錄製

這套語料是請中央社錄製的對話語音，從 2003 年六月開始到 2004 前五月，

內容為兩個以上（大部分是兩個）的播音員在廣播室閒聊。在聽取其錄音內容後再針對對話內容加以分類並給予主題。這套語料總共包括 49 個聲音檔，每個檔案錄音格式為標準的 Windows wav (Linear PCM)，合約十小時又四十二分鐘。大部分的對話者，說話速度適中，語調清晰，但是部份檔案音量偏小，播音員聲音含混，也有穿插笑聲、彼此搶話、也有在一邊吃東西一邊錄音等現象。由於是屬於自發性語音（並非照著稿子所唸），因此也常常出現各式各樣的語氣詞、停頓、口語不清、喘氣、搶話、笑聲、聲量忽大忽小、發音不清等現象，例如連續說"對對對對對"來贊同對方所講的話，發出"嗯"、"哦"、"那"、"哇"等語氣詞，在三人以上對話時，更時常會有搶話的現象。

所有分類的主題包括：居家、旅行與音樂、旅遊、運動、溫泉、新聞、算命、談狗狗、上台報告經驗談、工作、五專生活、吃點心談相親、狗狗命名、香港人、香辣蟹、租屋經驗、做菜、棒球、週年慶、閒話家常、節目討論、電視刻、漁人碼頭、養小狗、雜談、關東煮、什麼都有、好吃的郭老師、早餐吃什麼、晚餐吃什麼、買回來了、閒扯淡、閒話家常。

(4) 仿人機對話語音之錄製

這份語料是請中央社專業播音員模仿人與電腦對談的錄音，從 2003 年六月開始到 2004 年前五月，主要由一個人模仿自動語音對話系統中的電腦反應，另外一個人模仿詢問電腦的使用者，由中央社根據我們指定的網頁事先設計對話內容，所以算是仿自發性的語料。

對話（或問答）內容分為三大類，第一類是查詢跟國立台灣大學相關的問題（根據國立台灣大學的網頁<http://www.ntu.edu.tw>所設計），第二類是查詢旅遊的資訊（根據 CITN 台灣旅遊聯盟的網頁<http://travel.network.com.tw/>所設計），第三類是查詢餐廳資訊（根據咕嚕美食網的網頁<http://restaurant.goolu.com/>所設計）。

第一類（台大查詢）共有 207 則（每則約共有十二個問答句），一則平均約需 55 秒，207 則共需 2.45 小時；第二類（旅行查詢）共有 301 則（每則約共有八個問答句），301 則共需 3.49 小時；第三類（餐廳查詢）大約有 300 則，目前中央社尚在錄製當中。

截至目前為止，我們取得的仿人機對話錄音語料包括第一類（台大查詢）和第二類（旅行查詢）大約共有 6 小時，每個檔案錄音格式為標準的 Windows wav (Linear PCM)。以下分別為查詢台大資訊及旅遊資訊的兩則例子：

(a) 台大查詢部份：

- Q1: 台大網頁的字體很小，看了很吃力...
- A1: 喔，那...建議您瀏覽 30 分鐘要休息 10 分鐘...
- Q2: 這是眼睛保健的方法。除了這個解決方法，沒有別的了嗎？
- A2: 建議您朝改善視力的方向努力...
- Q3: 來不及了...已經過了青春期的視力也沒辦法恢復了...
- A3: 建議您用放大鏡試試看...
- Q4: 不能將網頁的字體放大嗎???
- A4: 這個問題問得好，可能是設計者視力 2.0，所以沒考慮到這個問題...
- Q5: 那...這個問題要向哪個單位反應？
- A5: 網站設計是由計算機及網路資訊中心程式設計組負責？
- Q6: 請問該如何聯絡他們？
- A6: 校內分機先撥 54，再撥 5052，校外請直撥 3366505

(b) 旅行查詢部份：

- Q1: 請問飯店的服務項目裡，所謂的“專車接送”我該如何告知飯店？
- A1: 您可於訂房成功後直接與飯店連繫，告知飯店您的接送地點與時間，屆時飯店即會派人前去接送。
- Q2: 我想指定住宿樓層，而且要面向海邊，如何處理？
- A2: 若有有關住宿樓層或住宿房間方位等個人化需求，請您於入住 check in 時向飯店提出要求，各飯店將會依您實際需求予以提供服務。
- Q3: 飯店所提供的免費餐點，是否有提供素食？
- A3: 有關住房後飲食的需求，請您於入住 check in 時向飯店提出要求，各飯店將會依您實際需求予以提供服務。
- Q4: 我若是遇天災等不可抗拒之因素，而影響到原本的訂房行程，該如何處理？
- A4: 所謂天災等不可抗拒之因素係以飯店所在地縣市政府頒布狀況為判定準則。若因天災等不可抗拒之因素，而欲取消訂房者，台灣旅盟建議您辦理保留訂房(保留期限依各飯店規定)，等天氣穩定後再行辦理預訂入住。若您仍決定取消訂房不辦理保留時，則台灣旅盟將視實際狀況，酌收新台幣 200 元電子商務服務手續費，另因各家飯店對因天災等不可抗拒之因素而取消訂房的規定仍有不同，台灣旅盟將與飯店做退房手續費之確認後，再以 email 或電話與您聯繫。

2.2 語料文字標註處理

我們將所收錄的三個主要語料—新聞語音、電台廣播語音及日常對話語音進行人工文字標註處理，處理方式分為兩種：(1) 新聞語音的標註是由受過訓練的專任助理進行較準確的文字標註，並做 cross checking，此部分工作由中央研究院資訊所王新民博士統籌規劃及負責推動；(2) 電台廣播語音及日常對話語音的標註是由受過簡單訓練的大學工讀生來做，由參與本計畫的各大學共同進行。以下簡述此兩部分，詳細的文字標註工作參見附件一、二。

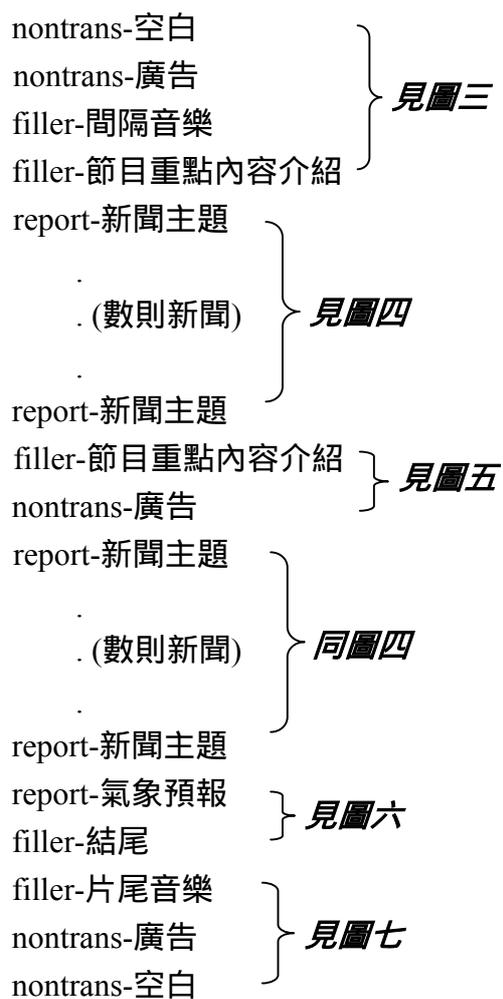
A. 新聞語音標註

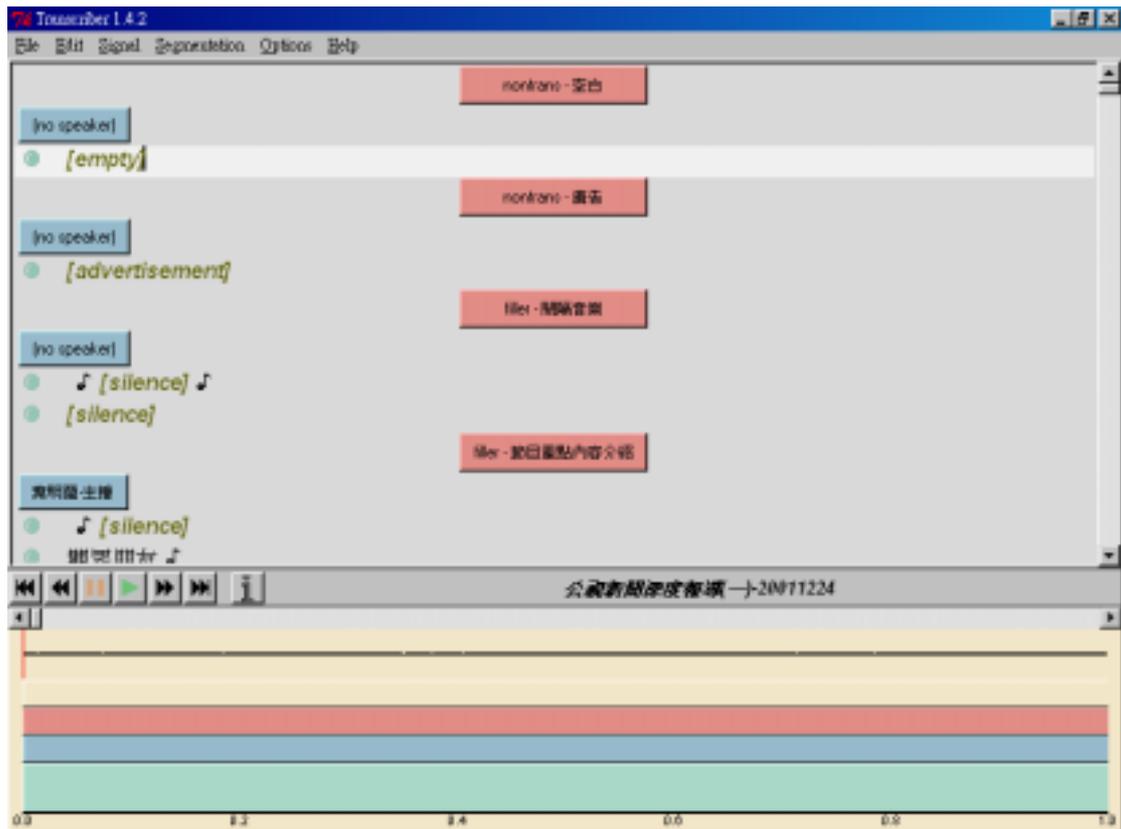
我們採用 LDC (Linguistic Data Consortium) 提供的 Transcriber 系統[1]來標註電視新聞錄音資料，請參考圖一。首先，將公視取回的 DAT(數位錄音帶)，經 USB 介面直接將錄音帶內的數位信號讀進 PC 內轉為格式為 44.1kHz 16bit stereo 的聲音檔 (windows PCM、.wav)，並燒錄於光碟中以便保存。然後，將檔案轉成標註使用的聲音檔，因考量檔案傳輸及讀取速度的問題，將原始的檔案，利用聲音編輯軟體 — CoolEdit 2000 將已轉為 windows PCM 的聲音檔進行格式轉換。轉換為 16kHz、16 bit、mono 後，為便利日後管理及利用，每週的公視新聞深度報導，每月的公視演講廳、客家新聞雜誌分別儲存於同一光碟中保存。

標註重點主要分為四大部分，分別為：

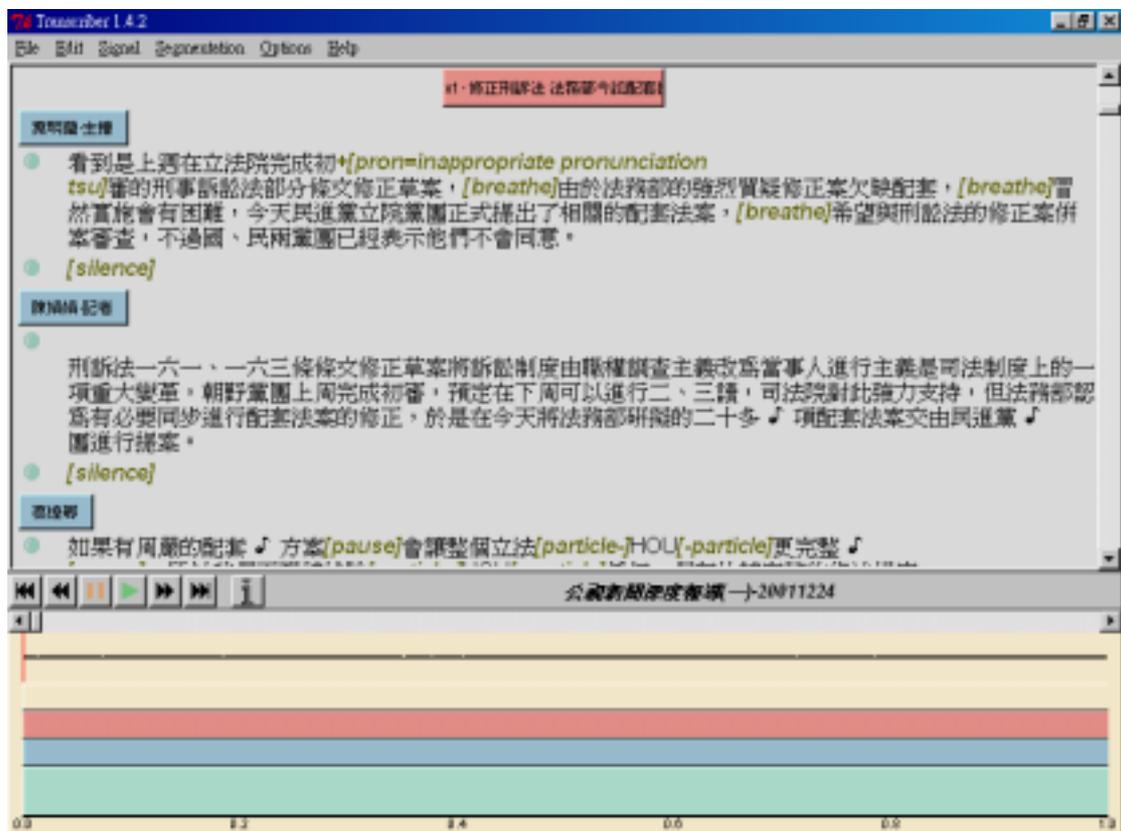
- ◆ 段落主題
- ◆ 說話者名稱
- ◆ 背景聲音
- ◆ 插入事件

『公視新聞深度報導』於 Transcriber 系統大致上之基本架構包含：

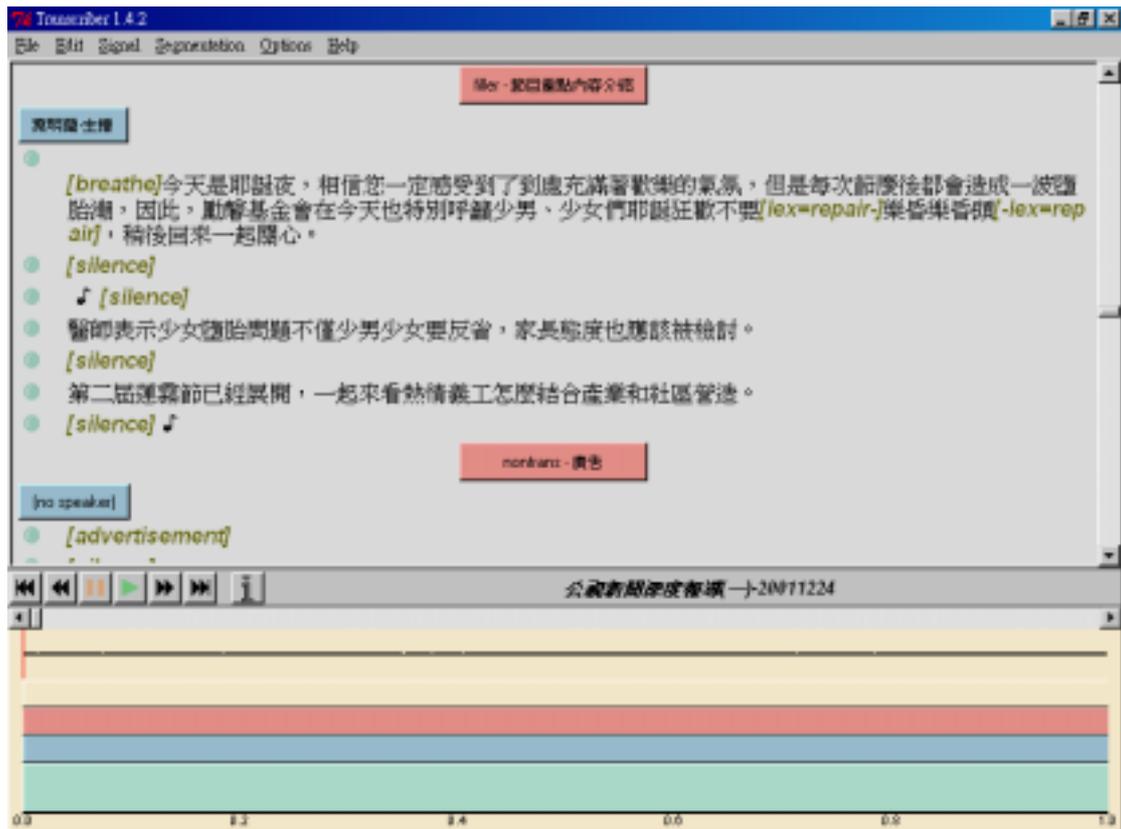




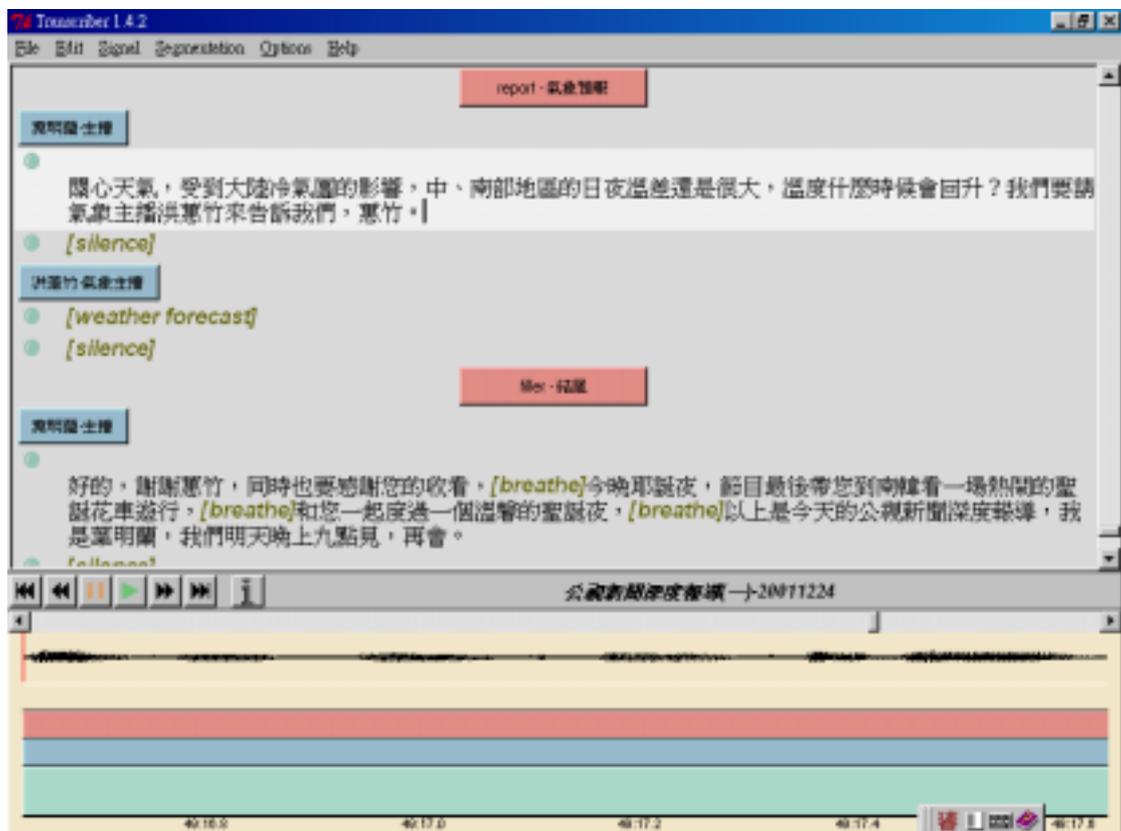
圖三



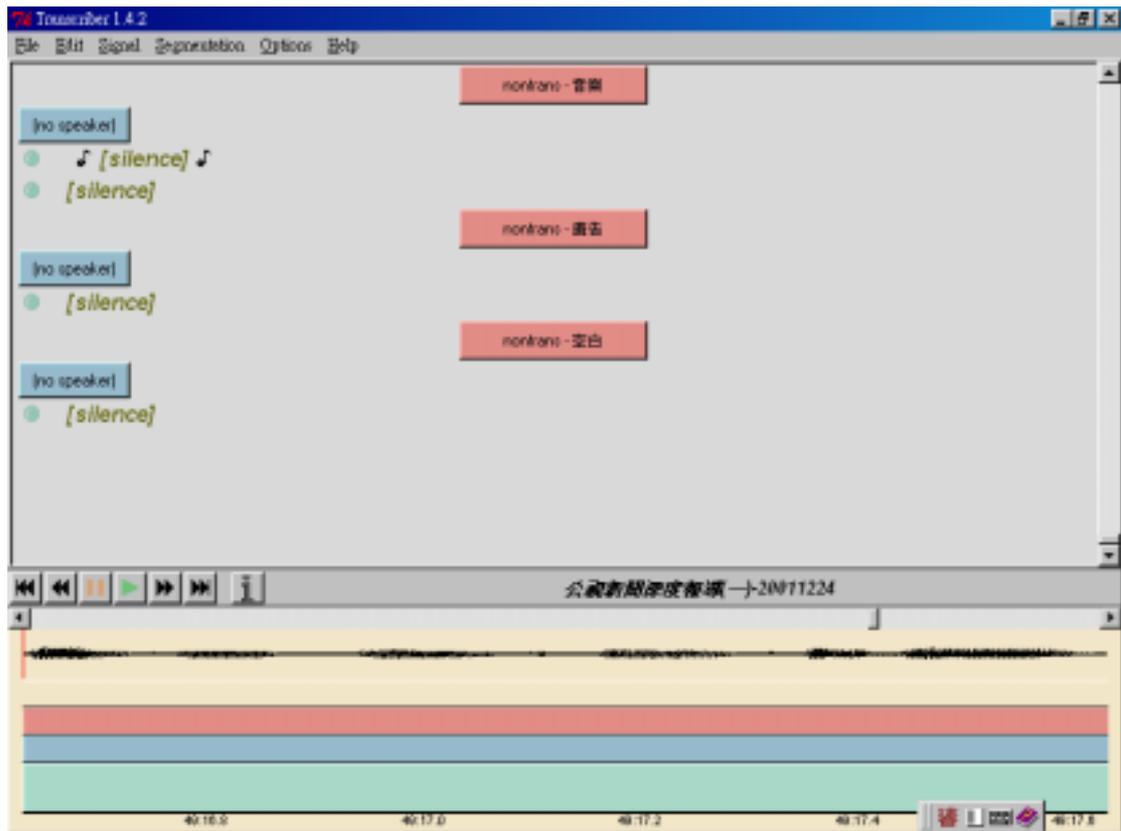
圖四



圖五



圖六



圖七

B. 電台廣播語音及日常對話語音標註

訪談內容以電台廣播方式呈現，內容大多為男、女主播採訪著名專業人士，剖析其相關領域的專業知識，有的檔案有獨立的主題，有些則是具連貫性主題的系列訪談。大部分的訪談錄音中，說話速度適中，語調清晰，但是部份檔案音量偏小，主播或受訪者聲音含混，也有穿插笑聲、彼此搶話等現象。

由於主播及被訪者的交談屬於自發性語音（並非照著稿子所唸），因此常常出現各式各樣的語氣詞、停頓、口語不清、喘氣、搶話、笑聲、聲量忽大忽小、發音不清等現象，例如連續說"對對對對對"來贊同對方所講的話，發出"嗯"、"哦"、"那"、"哇"等語氣詞；在多人訪談時，時常會有搶話的現象；而節目開始前通常會先來段音樂緩和氣氛或是訪談到一半大家想休息一下時，也會播放音樂。

我們請工讀生利用中研院推薦的語料標註軟體（Transcriber），從這些訪談的原始錄音資料聽取到的內容標記下來，除了將說話的文字內容以繁體中文標註之外，也仔細地標註每則訪談的語氣詞、停頓、口語不清、喘氣、搶話、笑聲、聲量忽大忽小、發音不清等現象，使我們日後在自發性語音辨識研究方面能有豐富的研究資料。

除了將自發性語音中會發生的種種現象也清楚標記之外，我們也請工讀生將這些語音及標記作進一步的分析整理，根據 transcriber 的標記，將每則訪談的語音檔切割成一句話一個語音檔，然後以句子為單位在資料庫(MS SQL)中分別記錄其標記特性。因為當我們想要找所有有搶話特性的句子時，如果靠人一句一句地慢慢找的話並不可行，所以我們將語氣詞、停頓、口語不清、喘氣、搶話、笑聲、聲量忽大忽小、發音不清等這些特性在資料表中分別分配一個欄位，有此特性這欄位就標成 true，沒有的話則標成 false，例如某句子有夾雜笑聲的話，就將這句子的笑聲欄位設成 true。這些欄位建立好後，我們只要利用 SQL 的 query language 就可以在短短幾分甚至幾秒中查到我們想要的檔案或語料長度等。

第三章 結果與討論

經過三年的努力，我們共收錄了包括新聞語音、電台廣播語音、日常對話語音、及仿人機對話語音四種華語自發性語音語料，並對其中的新聞語音、電台廣播語音、及日常對話語音三個語料進行人工文字標註處理，成果說明如下：

3.1 語料收錄及標註結果

A. 新聞語音處理

第一年度預計完成的第一階段 40 小時的語料庫文字標註處理已於 91 年 7 月底如期完成；第二年度預計完成的第二階段 80 小時的語料庫文字標註處理也已於 92 年 7 月底如期完成；本年度則完成 78 小時的語料庫文字標註處理。

除進行語料標註工作之外，我們也將前二年度完成的 120 小時語料庫進行一次的完整修訂工作，主要是將語料庫中原標註不一致處訂定統一標準，盡量達到語料標註的一致性，另外，我們也針對此一第一年度 40 小時語料庫進行初步的統計分析，並寫成一篇會議論文發表於 2003 IEEE&ISCA Workshop on Spontaneous Speech Processing and Recognition [2]。第一年度的 40 小時語料共包括 40 集新聞錄音及其標註檔案，每五集儲存在一片光碟，共有 8 片光碟，第二年度的 80 小時語料庫則以兩片 DVD+R 儲存，此二語料庫均已送交本計畫其他共同執行單位測試中，相信很快可以授權學術單位或產業界使用。

B. IC 電台語料處理

IC 電台語料大部分都已處理，共有 70616 字，合 8 小時 36 分 12 秒，下表為處理結果：

編號	節目名稱	檔名	總長度	中文音節數	Speaker 數
1	IC 三朵花	20020426_1000	59 分 4 秒	3044	2
2	IC 三朵花	20020426_1100	55 分 2 秒	5903	4

3	IC 晚報	20020425_1800	59 分 36 秒	10540	10
4	IC 晚報	20020425_1900	60 分 10 秒	10060	5
5	歡喜一家親	20030405_1	59 分 13 秒	7081	3
6	歡喜一家親	20030405_2	59 分 3 秒	5053	3
7	愛上新竹	20030405_3	60 分 4 秒	9887	3
8	愛上新竹	20030412_3	52 分 50 秒	9015	3

註 1：以上檔案未將廣告、音樂...等非語音去除的部份

編號	節目名稱	檔 名	總長度	中文字數	Speaker 數
9	歡喜一家親	20030412_1_1	1 分 41 秒	452	2
		20030412_1_2	3 分 27 秒	494	4
		20030412_1_3	3 分 54 秒	474	3
		20030412_1_4	4 分 02 秒	977	3
		20030412_1_5	3 分 42 秒	1013	3
		20030412_1_6	3 分 44 秒	879	3
		20030412_1_7	1 分 11 秒	288	2
10	歡喜一家親	20030412_2_1	5 分 49 秒	1184	3
		20030412_2_2	5 分	1201	3
		20030412_2_6	50 秒	226	2
		20030412_2_7	1 分 34 秒	235	2
		20030412_2_8	1 分	191	2
11	心靈故鄉	20030804_1_1	6 分 51 秒	1421	2
		20030804_1_2	8 分 25 秒	1908	2
合計			3 時 36 分 12 秒	70616	

註 2：以上檔案已去除非語音音段，並將節目切成多個小段

C. 中央社對話語料處理

中央社對話語料只處理部分檔案，共有 50566 字，合 3 小時 23 分 30 秒，下表為處理結果：

編號	節目名稱/檔名	總長度	音節數	Speaker 數
1	居家 2	10 分 10 秒	2965	2
2	居家 3	9 分 55 秒	2911	2
3	居家 4	10 分 58 秒	3145	2
4	新聞	10 分 36 秒	3271	2
5	溫泉	35 分 49 秒	9654	2
6	運動	10 分 00 秒	2871	2
7	算命 5	10 分 15 秒	2512	2
8	算命 6	10 分 28 秒	2975	2
9	談狗狗	10 分 37 秒	2598	2
10	旅行與音樂 1	17 分 16 秒	3091	2
11	旅行與音樂 2	15 分 41 秒	3123	2
12	旅行與音樂 3	15 分 08 秒	2332	2
13	旅行與音樂 4	15 分 09 秒	2736	2
14	旅遊 1	10 分 13 秒	3115	2
15	遊遊 2	11 分 15 秒	3267	2
合計		3 時 23 分 30 秒	50566	

D. 中央社仿人機對話語料處理

中央社仿人機對話語料是請中央社專業播音員模仿人與電腦對談的錄音，主要由一個人模仿自動語音對話系統中的電腦反應，另外一個人模仿詢問電腦的使用者，由中央社根據我們指定的網頁事先設計對話內容，所以算是仿自發性的語料。

依對話（或問答）內容此語料可分為三大類，第一類是查詢跟國立台灣大學相關的問題（根據國立台灣大學的網頁<http://www.ntu.edu.tw>所設計），第二類是查詢旅遊的資訊（根據 CITN 台灣旅遊聯盟的網頁<http://travel.network.com.tw/>所設計），第三類是查詢餐廳資訊（根據咕嚕美食網的網頁<http://restaurant.goolu.com/>所設計）。我們共錄製了第一類（台大查詢）有 207 則（每則約有十二個問答句），一則平均約需 55 秒，207 則共需 2.45 小時；第二類（旅行查詢）共有 301 則（每則約共有八個問答句），301 則共需 3.49 小時；第三類（餐廳查詢）大約有 300 則。此語料目前只做初步處理，尚未以人工做文字標註處理。

第四章 結論

本計畫共收集了四種華語自發性語音資料，包括新聞語音、電台廣播語音、日常對話語音、及仿人機對話語音，並對其中的新聞語音、電台廣播語音、及日常對話語音三個語料進行人工文字標註處理，完成了 198 小時的廣播新聞語音處理；8 小時 36 分 12 秒的電台廣播語音處理，含 70616 字；及 3 小時 23 分 30 秒的日常對話語音處理，含 50566 字。

本計畫有許多的後續工作，首先已完成文字標注的語料庫需先試用，進行實際的語音辨認模型訓練及測試，以了解資料的正確性，並逐漸建立各種自發性華語語音辨認的 benchmark，以作為未來進一步研究的 reference；接著我們需要對語料庫做進一步處理，將各種自發性語音現象進行標註，例如音節間嚴重耦合 (coarticulation)，另外對語者發音的情緒、對話的意圖分類等標註也需進一步探討；最後，語料庫的擴充也是值得進行的工作，自發性語料的種類相當多，我們只收集幾類重要的語料，未來應再收集其他種類的語料，例如演講語料。

自發性語音資料庫的收集及文字標註處理是一件十分繁雜的工作，耗費相當多的人力，幸賴參與本計畫的各單位師生同仁共同努力，貢獻心力與時間，才得以順利完成。未來這些語料庫將提供國內外學術界使用，將成為華語語音處理領域十分重要的研究素材。

致 謝

本計畫受許多單位及個人的協助才得以順利完成，包括：國科會的經費贊助、公共電視台的錄音及語料授權、中央社的錄音及語料授權、IC 電台的同意錄音、中研院曾淑娟博士的文字標註指導、以及各參與單位的人力物力投入，在此一併致謝。

參考文獻

- [1] Barras, E. Geoffrois, Z. B. Wu, M. Liberman, “Transcriber: Development and Use of S tool for Assisting Speech Corpora Production,” *Speech Communication*, 33, pp. 5-22, 2001.
- [2] Hsin-min Wang, "MATBN 2002: A Mandarin Chinese broadcast news corpus," in *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, Tokyo, April 2003.
- [3] Gary Cook, Tony Robinson, and James Christie, “Real-time Recognition of Broadcast News,” *ICSLP98*.
- [4] Ha-Jin Yu, et al, “Automatic Recognition of Korean Broadcast News Speech,” *ICSLP98*.
- [5] P. C Woodland, et al, “Experiments in Broadcast News Transcription,” *ICASSP98*, pp. 909-912.
- [6] Gary Cook and Tony Robinson “Transcribing Broadcast News with the 1997 ABBOT System,” *ICASSP98*, pp. 917-920.
- [7] Petra Geutner, Michael Finke, and Peter Scheytt, “Adaptive Vocabularies for Transcribing Multilingual Broadcast News,” *ICASSP98*, pp. 925-028.
- [8] Matsuoka T., Taguchi Y., Ohtsuki K., Furui S., and Shirai K., “Toward Automatic Transcription of Japanese Broadcast News,” *Eurospeech97*, pp. 915-918.
- [9] T. Imai, A. Kobayashi, S. Sato, H. Tanaka, and A. Ando. “Progressive 2-pass Decoder for Real-time Broadcast News Captioning,” *ICASSP2000*.
- [10] TREC Web site: <http://trec.nist.gov/>
- [11] TDT3 Web site: <http://www.itl.nist.gov/iaui/894.01/tdt3/announce.htm>
- [12] LDC Web site: <http://www ldc.upenn.edu/>
- [13] David C. Abberley, Steve J. Renals, and Gary Cook, “Retrieval of Broadcast News Document with The THISL System,” *ICASSP98*, pp. 3781-3784.
- [14] Philip Clarkson and Tony Robinson, “The Applicability of Adaptive Language Modeling for the Broadcast News Task,” *ICSLP98*.
- [15] Thomas Hain and P. C. Woodland, “Segmentation and Classification of Broadcast News Audio,” *ICSLP98*.
- [16] Ara Samouelian, Jordi Robert-Ribes, and Mike Plumpe, “Speech, Silence, Music and Noise Classification of TV Broadcast Material,” *ICSLP98*.

- [17] Jean-Luc Gauvain, Lori F. Lamel, and Gilles Adda, "Partitioning and Transcription of Broadcast News Data," ICSLP98.
- [18] Alexander G. Hauptmann and Michael J. Witbrock, "Story Segmentation and Detection of Commercials in Broadcast News Video," ADL98.
- [19] Ohtsuki K., Matsuoka T., Matsunaga S., and Furui S., "Topic extraction with Multiple Topic-words in Broadcast News Speech," ICASSP98, pp. 329-332.
- [20] John Choi, et al, "SCAN – Speech Content Based Audio Navigator: The System Overview," ICSLP98.
- [21] Smith M., and Kanade T., "Video Skimming and Characterization through the combination of Image and language Understanding Techniques," Computer Vision and Pattern Recognition Conference, pp. 775-781, 1997.
- [22] Jean-Luc Gauvain, Lori Lamel, Yannick de Kercadio, and Gills Adda, "Transcription and Indexation of Broadcast Data," ICASSP2000.

附件

- 一、電視新聞錄音資料處理技術報告
- 二、電視節目錄音資料及對話語音資料處理技術報告

附件一：電視新聞錄音資料處理技術報告

目錄：

1. 原始資料
 - 1.1 來源
 - 1.2 節目
 - 1.3 錄音/錄影配備
2. 資料保存
 - 2.1 聲音資料
 - 2.2 影音資料
3. 標註
 - 3.1 基本架構
 - 3.2 語音資料標誌
4. 實際操作問與答

第一章 原始資料

1.1 來源: 公共電視公司

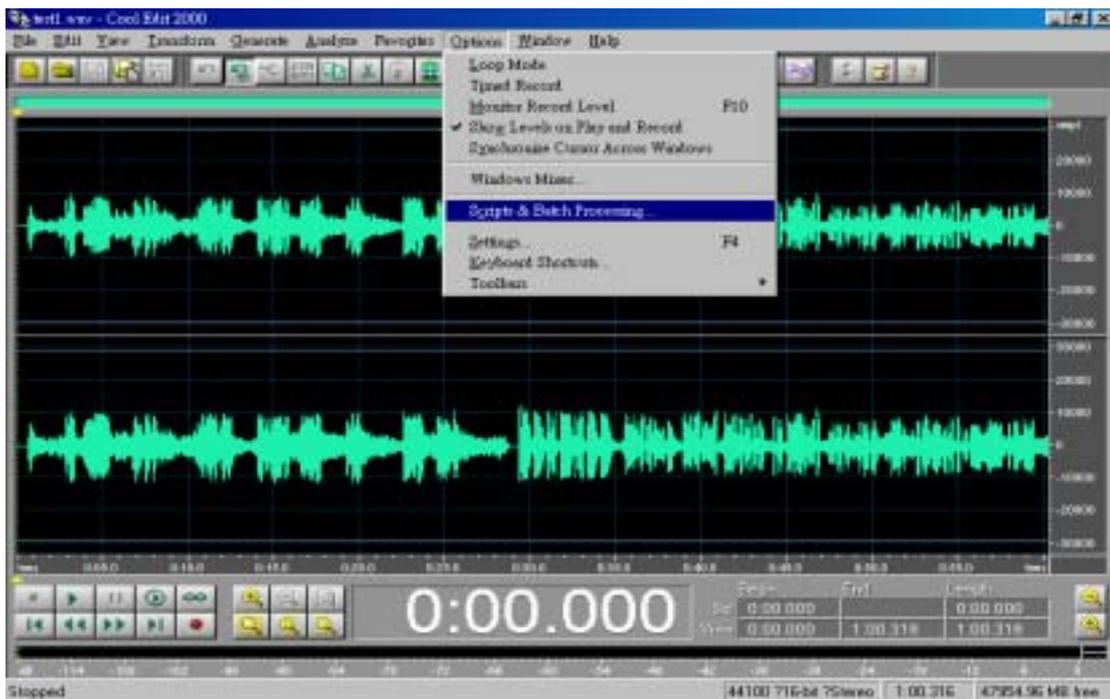
1.2 節目:

自 11 月 7 日起, 錄影/錄音下列節目

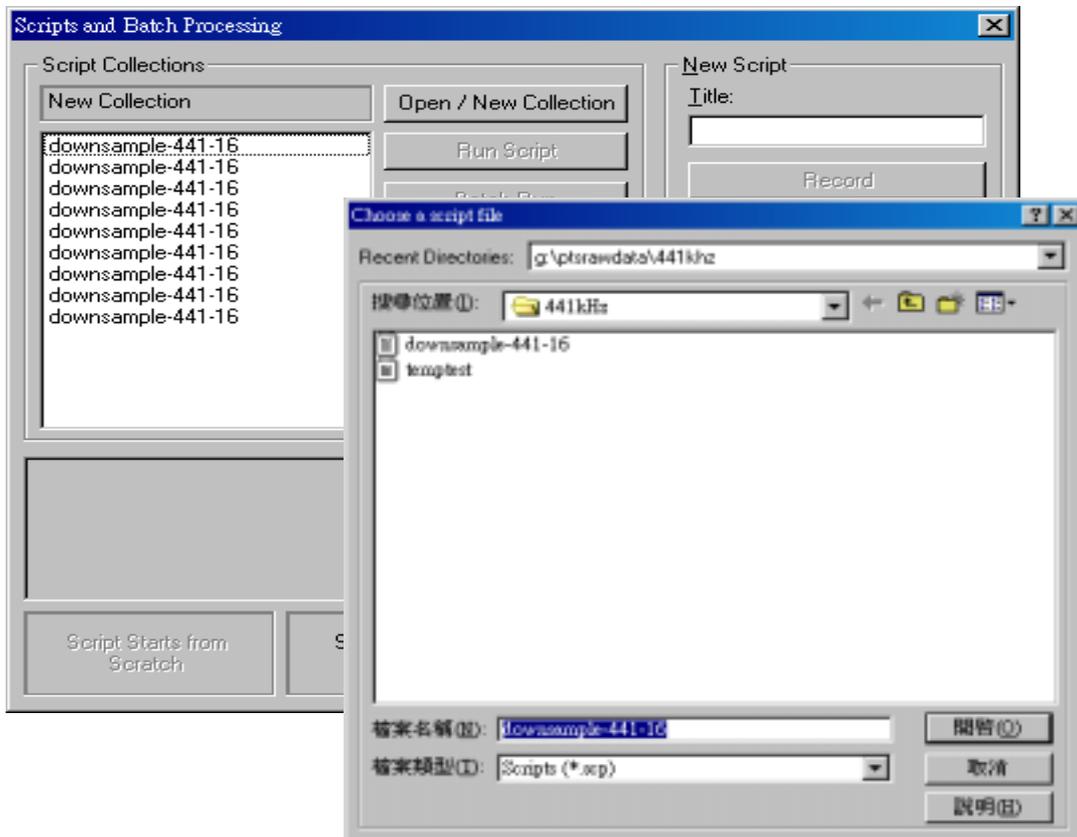
- I. 公視新聞深度報導 (週一 ~ 週五 21:00~22:00)
- II. 公視演講廳 (週六 15:00~16:00)
- III. 客家新聞雜誌 (週日 21:00~22:00)

1.3 錄音/錄影配備

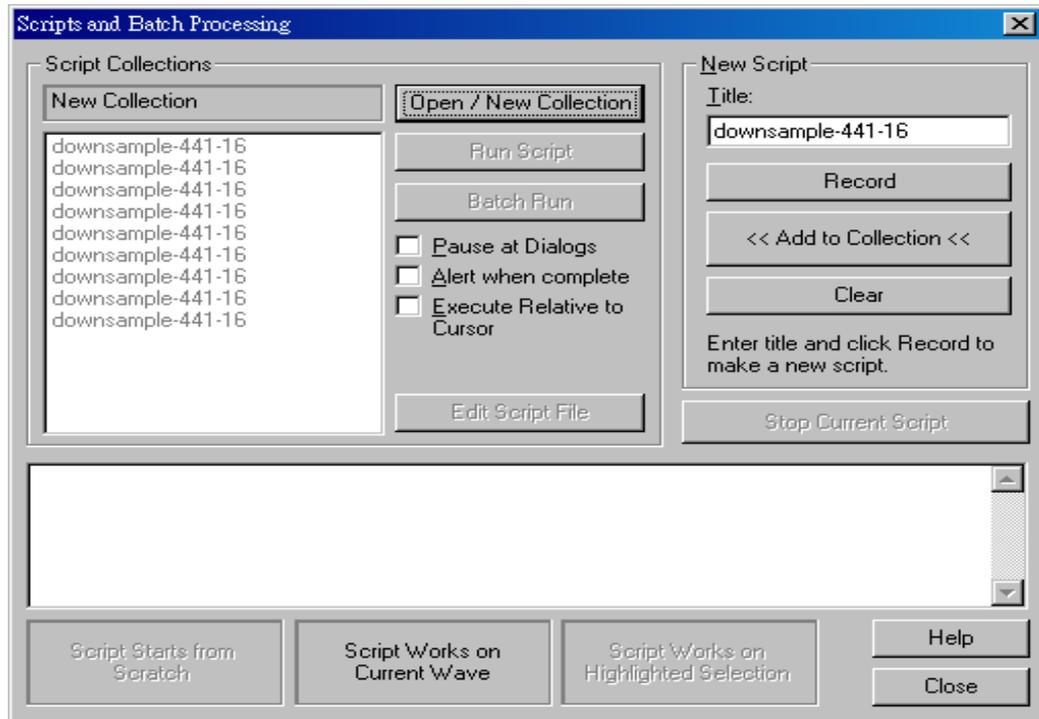
- 錄音採 TASCAM DA-40 DAT 錄音座, 經由主控台在新聞播放時利用 AES/EBU 平衡式類比輸入同步錄音。
- 錄影採 SONY SLV-ED88 錄放影機, 利用一般 RCA 接頭同步錄影, 錄影帶採用 TDK HS-160 型號。
- 錄音/錄影格式
 - ✓ DAT tape: 格式: 44.1kHz、16bit、stereo
 - ✓ VHS tape:
 - *DAT 及 VHS 待轉錄完成需重複使用。(DAT 錄音帶 VHS 錄影帶各有 60 捲, 28 捲)



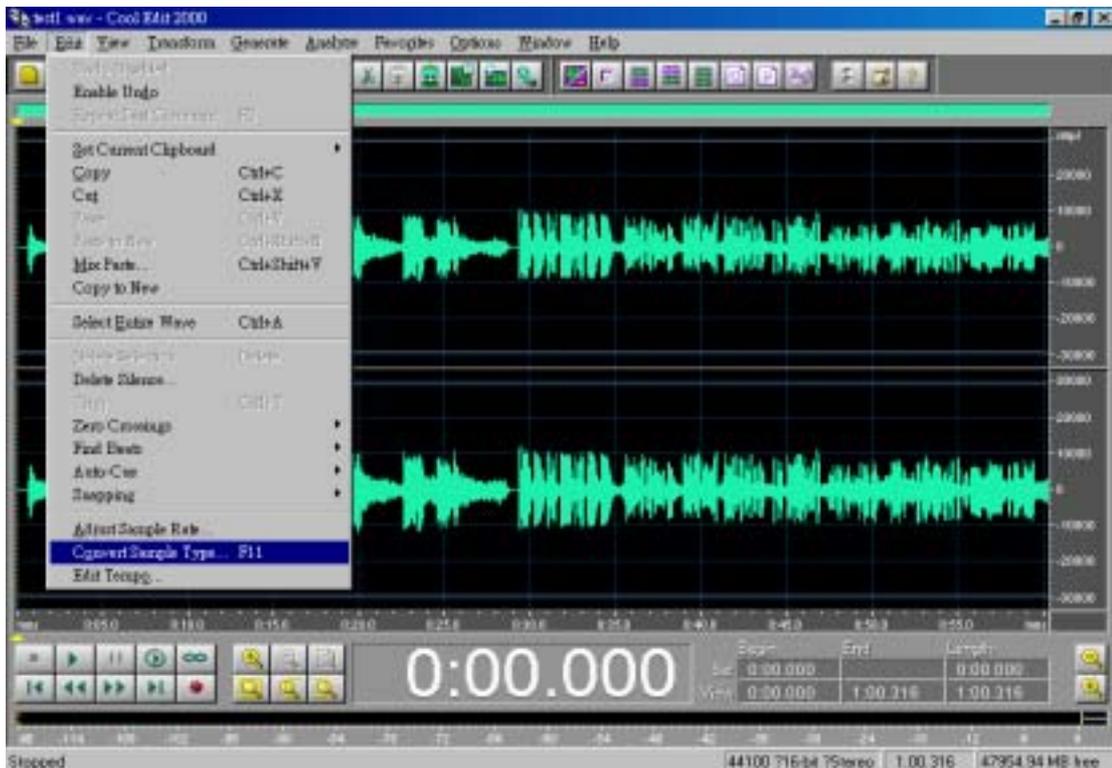
- D. 在『Scripts & Batch Processing』視窗中選擇『Open/New Collection』
 (在『Choose a Script file』視窗中開啟『downsample-441-16』檔案)



- E. 將『 Scripts & Batch Processing 』視窗中『 New Script 』之『 Title 』項目設定為『 downsample-441-16 』(按下『 Record 』鍵)

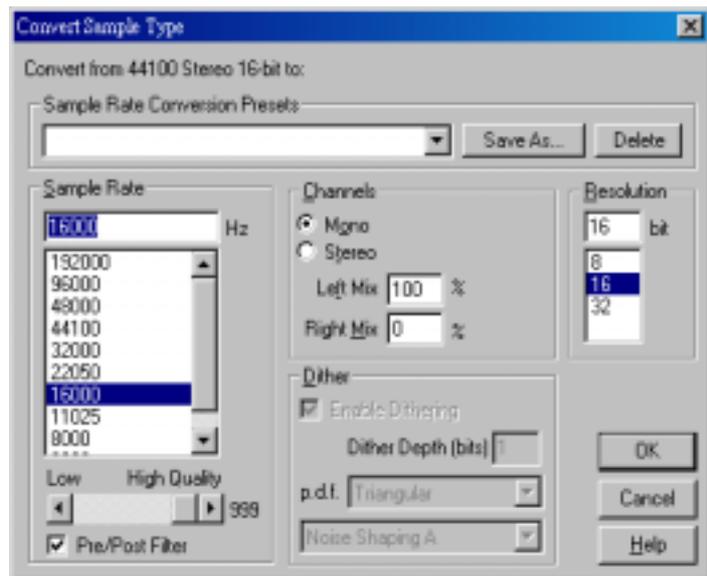


- F. 點選『 Edit 』(選擇『 Convert Sample Type 』選項(出現『 Convert Sample Type 』視窗



G. 在 『 Convert Sample Type 』 視窗中，將其內容設定如下：

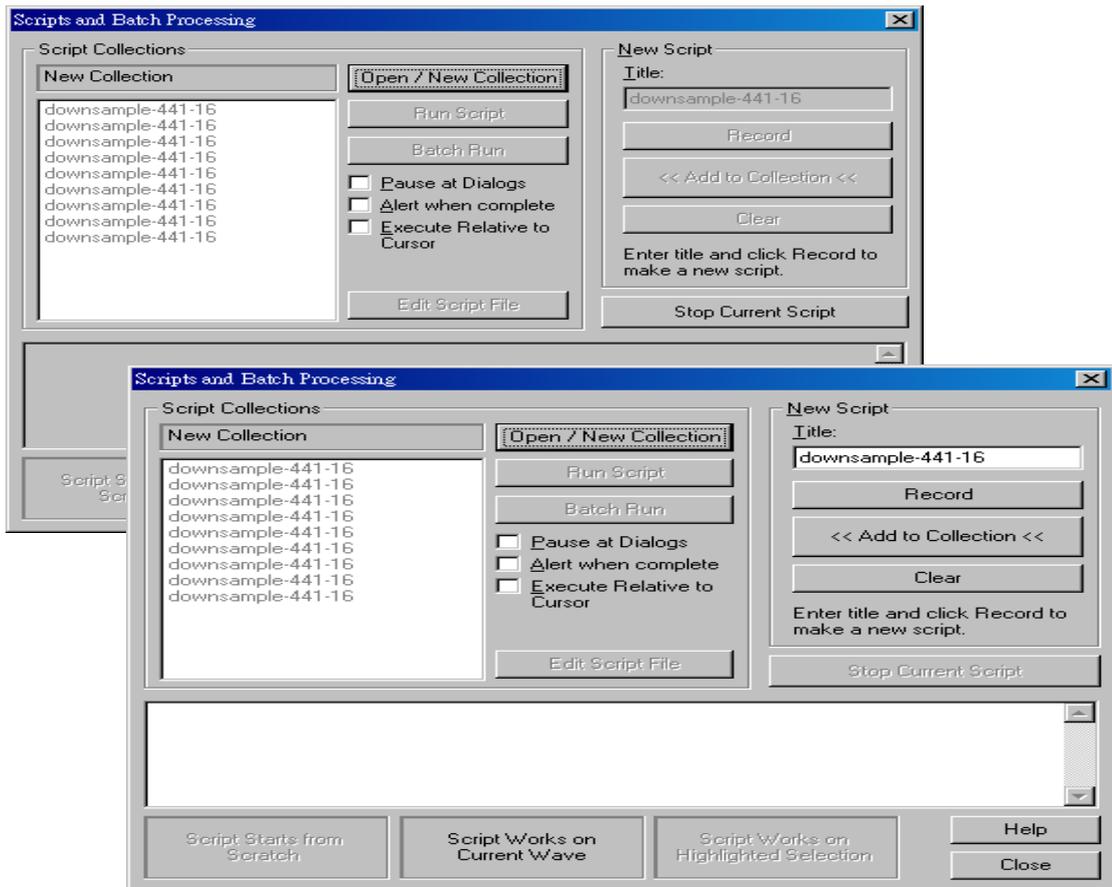
Sample Rate : 16000
High Quality : 999
Pre/Post Filter
Channels : Mono
Left Mix : 100%
Right Mix : 0%
Resolution : 16bit



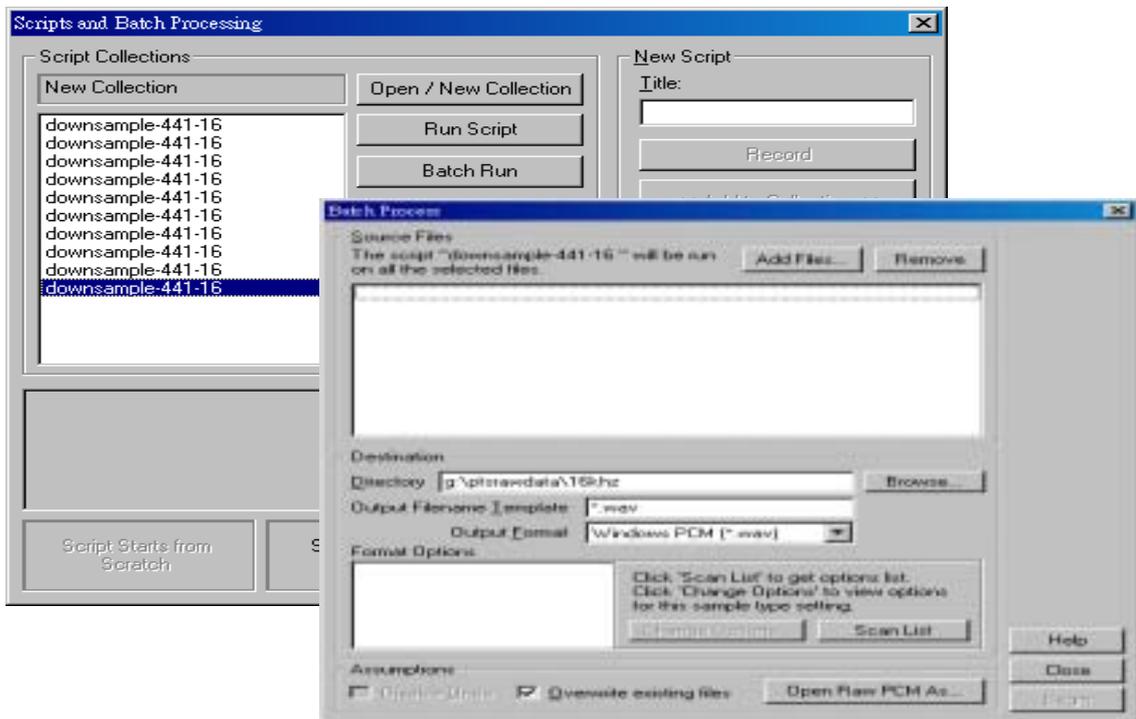
H. 點選 『 Option 』 → 選擇 『 Scripts & Batch Processing 』 選項



I. 在 『 Scripts & Batch Processing 』 視窗中選擇 『 Stop Current Script 』
→ 選擇 『 <<Add to Collection<< 』 項目



J. 點選『Batch Run』→出現『Batch Process』視窗→選取檔案(可同時選取多個檔案)→Add file

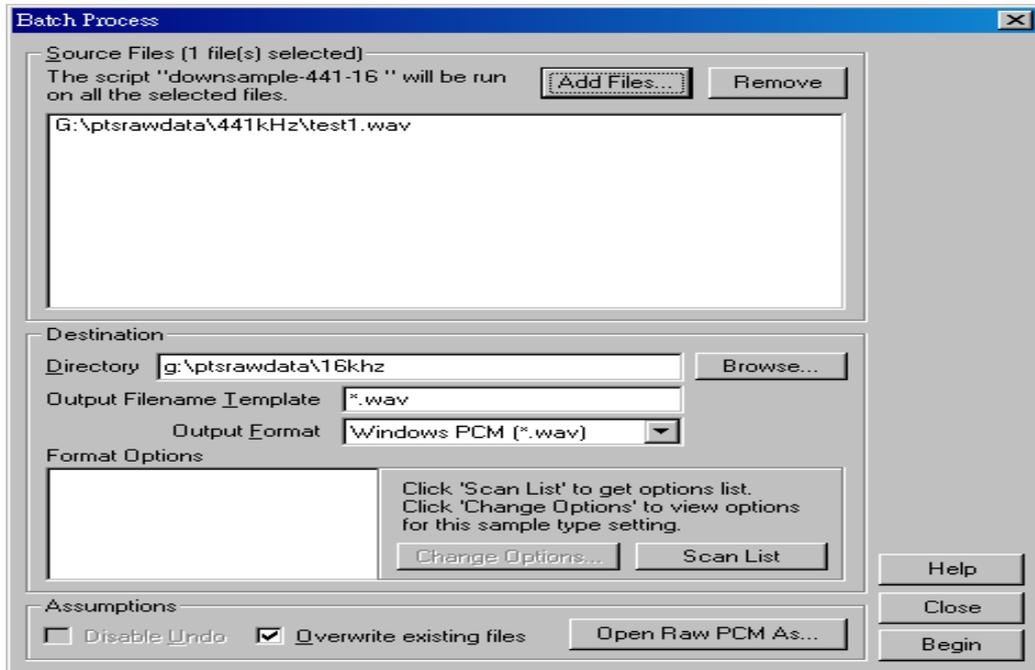


K. 將『Destination』設定如下：

Directory : g:\ptsrawdata\16khz

Output Filename Template : *.wav

Output Formate : Windows PCM [*.wav]



L. 點選『Begin』鍵即開始轉檔

附註:

- 由公視發送的節目，經詢問後了解，其左右聲道為同一輸入，但兩者的振幅大小略有變化，以做為立體聲兩聲道之區別。
- DAT 錄音帶轉錄時由於人為操控機器的因素影響，會有不同步的問題，即轉錄信號一開始處會有數個 sample 會是 0 值，後面也會有幾個 sample 是 0，但此情形並不影響真正記錄在 DAT 錄音帶內的數位信號。

2.2 影音資料

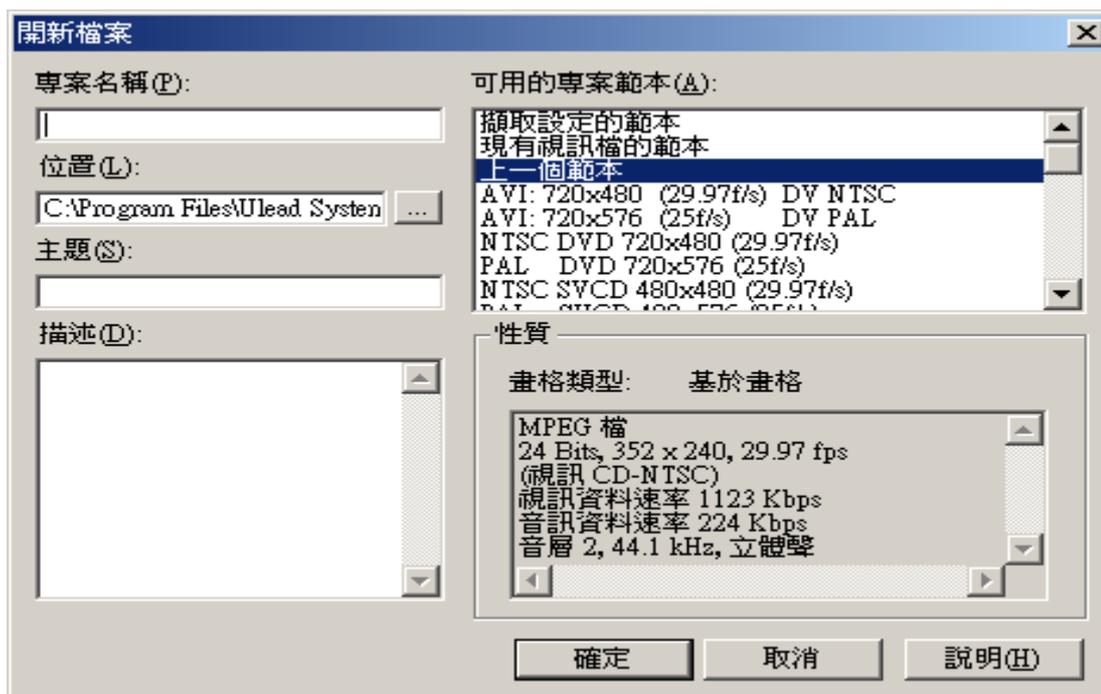
保存:

- I. 公視取回的 VHS 錄影帶，經由 UPMOST 301BTR 類比影像擷取卡，擷取 avi 格式的影像，並由影像編輯軟體 – 會聲會影(友立出品)即時壓縮成 MPEG1 格式，使用會聲會影進行格式轉換的詳細步驟如下：

A. 開新檔案



B. 設定『專案名稱』及『存檔位置』



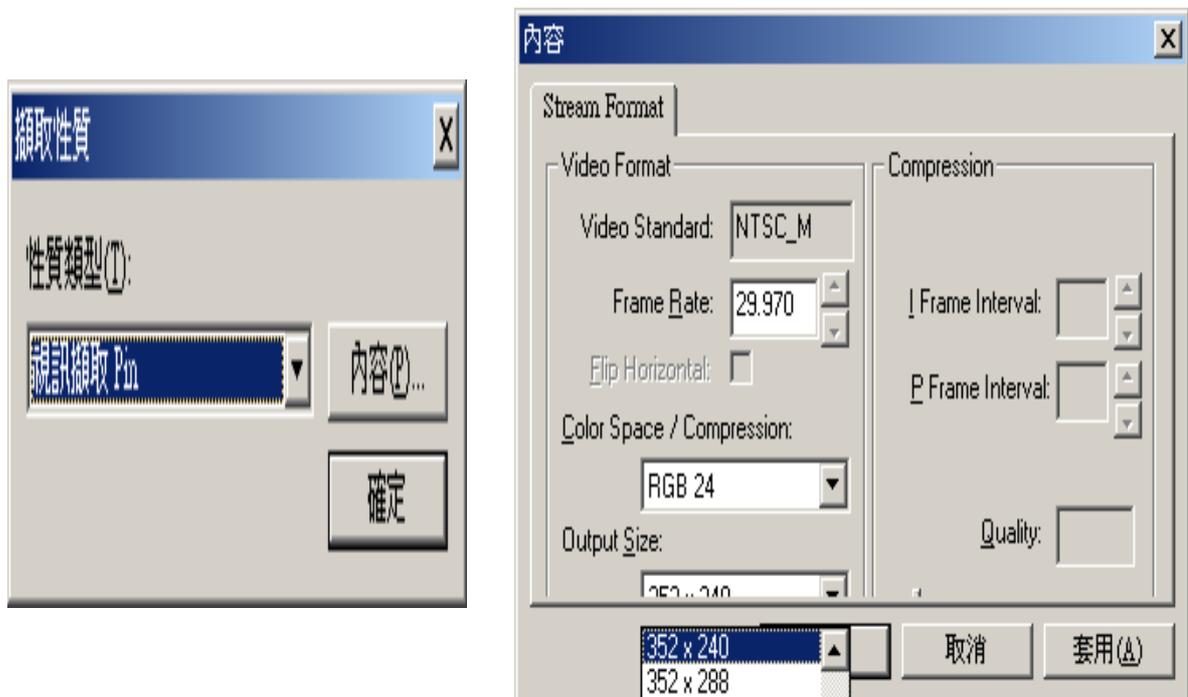
C. 打開『視訊』功能表，點選『視訊性質』



D. 在『擷取性質』視窗之『性質類型』中選擇『視訊擷取 Pin』項目，將其內容設定為：

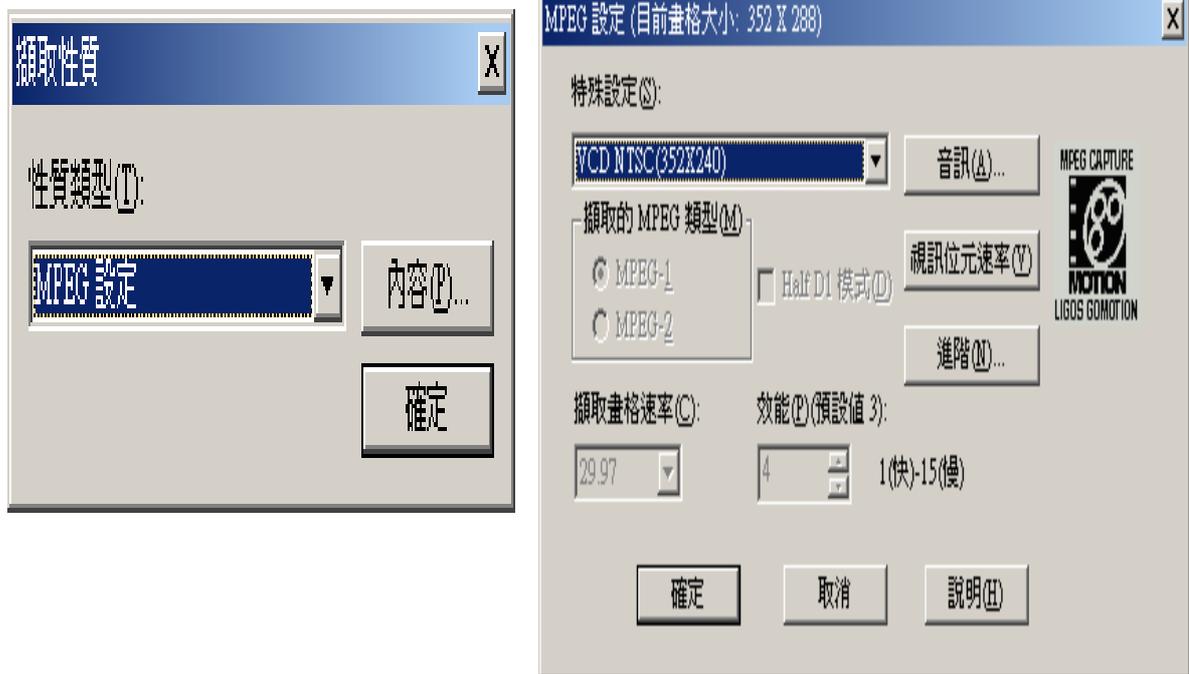
Color Space/Compression : RGB24

Output Size : 352 x 240



E. 在『擷取性質』視窗之『性質類型』中選擇『MPEG 設定』項目，

將其內容之『特殊設定』設定為：VCD NTSC(352x240)



- F. 點選『擷取視訊』，錄影機在放影的狀態下，於需開始轉檔的地方按下『會聲會影』視窗中之『確定』鍵，即可開始執行轉檔，中途暫停或轉檔完成則使用鍵盤左上角之『ESC』鍵即可停止轉檔動作



II. 90 年 11 月 7 日起至 91 年 1 月 6 日的錄影帶是由下列設定擷取成 MPEG1 後輸出成 VCD 格式再燒錄至 CD-ROM 以影音光碟形式保存。

專案檔範本屬性:

基於畫格

MPEG 檔

24 Bits、352 x 240、29.97 fps

(MPEG-1)

視訊資料速率 1123 Kbps

音訊資料速率 256 Kbps

音層 2、44.1 kHz、立體聲

擷取結束後於完成→ 輸出格式中選擇 **CD-NTSC 352 x 240**

MPEG 檔

24 Bits、352 x 240、29.97 fps

(視訊 CD-NTSC)

視訊資料速率 1123 Kbps

音訊資料速率 224 Kbps

音層 2、44.1 kHz、立體聲

III. 91 年 1 月 7 日起至 91 年 1 月 13 日為止(指節目播出日期),擷取格式變動,可省去輸出格式為 VCD 時所額外花費的時間,(約數十分鐘至半小時不等,視電腦執行成效而定),擷取下來的 MPEG1 檔案可直接由燒錄軟體轉成 VCD 格式,燒錄成影音光碟形式保存。擷取設定如下:

專案檔範本屬性

基於畫格

MPEG 檔

24 Bits、352 x 240、29.97 fps

(視訊 CD-NTSC)

視訊資料速率 1123 Kbps

音訊資料速率 224 Kbps

音層 2、44.1 kHz、立體聲

IV. 91 年 1 月 14 日起,由於考量日後檔案使用便利性,捨 VCD 檔案格儲存原始資料,而直接保存擷取下來的 MPEG1 檔案形式,自此日起,將不再有 VCD 影音光碟的原始資料保存,取代的保存資料為一般資料類型的 CD-ROM 格式,儲存資料為 MPEG1.變更儲存類型的理由如下:

一般 VCD 所儲存的影片是使用 .DAT 的檔案格式,雖然同樣是屬於 MPEG 的方式壓縮,但是和一般影像處理軟體所能處理的影片檔。MPG

還是有所不同。所以想要把 VCD 裡面的影片抓出來編輯，還是需要一個轉換的動作。經過測試，將 VCD 中的 DAT 檔轉成 MPEG 格式，若利用免費軟體 VCD Gear 來進行轉換，56 分鐘的 VCD 需時 7~8 分鐘的時間來轉換，為此，我們選擇將保存類型更改為 MPEG 格式，雖然此儲存格式將無法在家用 VCD player 上播放，但考慮到日後處理此檔案的環境多為個人電腦，MPEG 格式即可支援，故不另行備份 VCD 格式的影音光碟。

附註：

- 會聲會影在擷取影像時，常會有設定值跑掉的情形出現，此時會有錯誤訊息出現，需回到擷取設定中，將設定值調回目前設定的格式。

第三章 標註

電視新聞錄音資料之處理是採用 Transcriber 系統,其標註重點主要分為四大部分,分別為:

- A. 段落主題
- B. 說話者名稱
- C. 背景聲音
- D. 插入事件

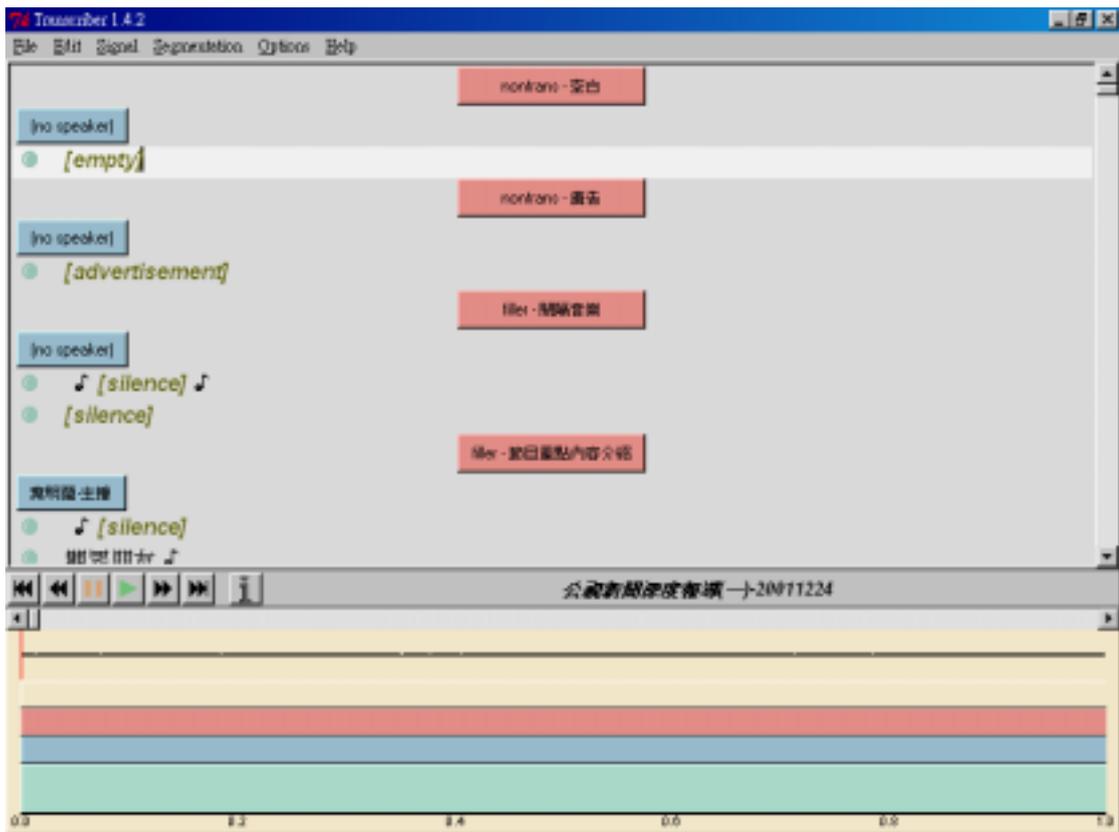
以下將針對上述四項說明其使用方法、標註方法以及標註上需注意之重點。

3.1 基本架構

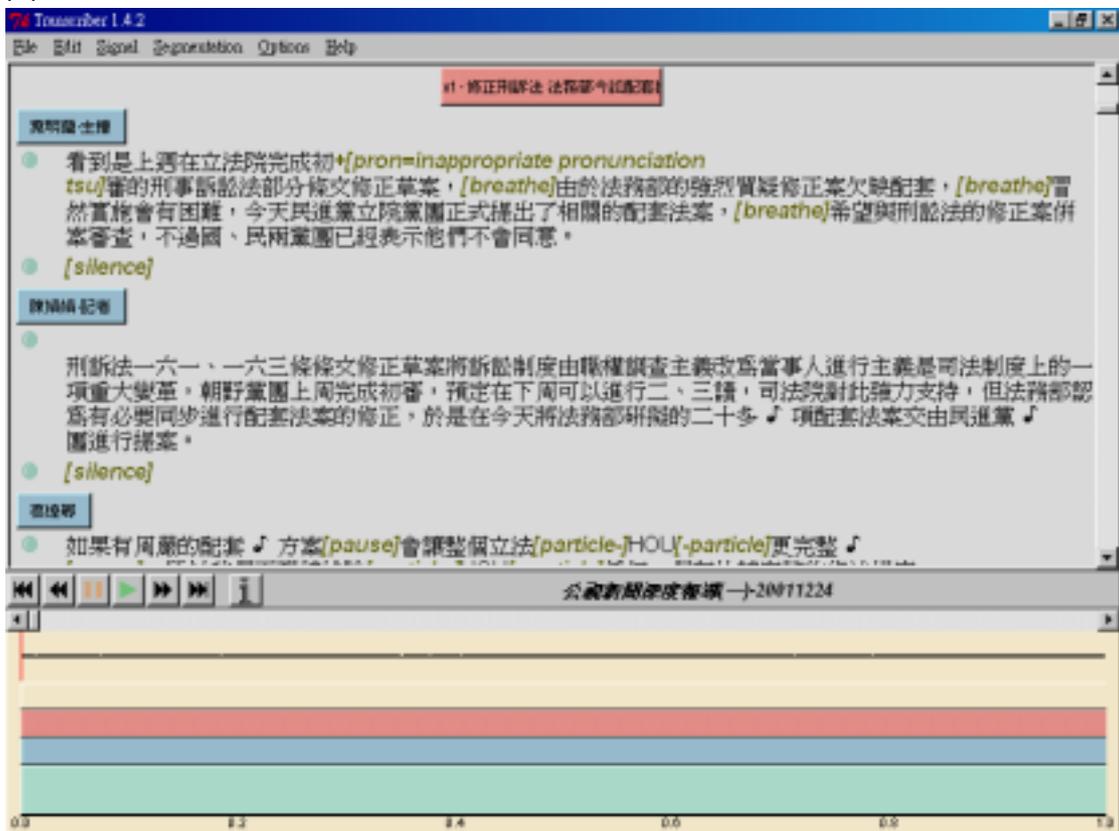
『公視新聞深度報導』於 Transcriber 系統大致上之基本架構包含:



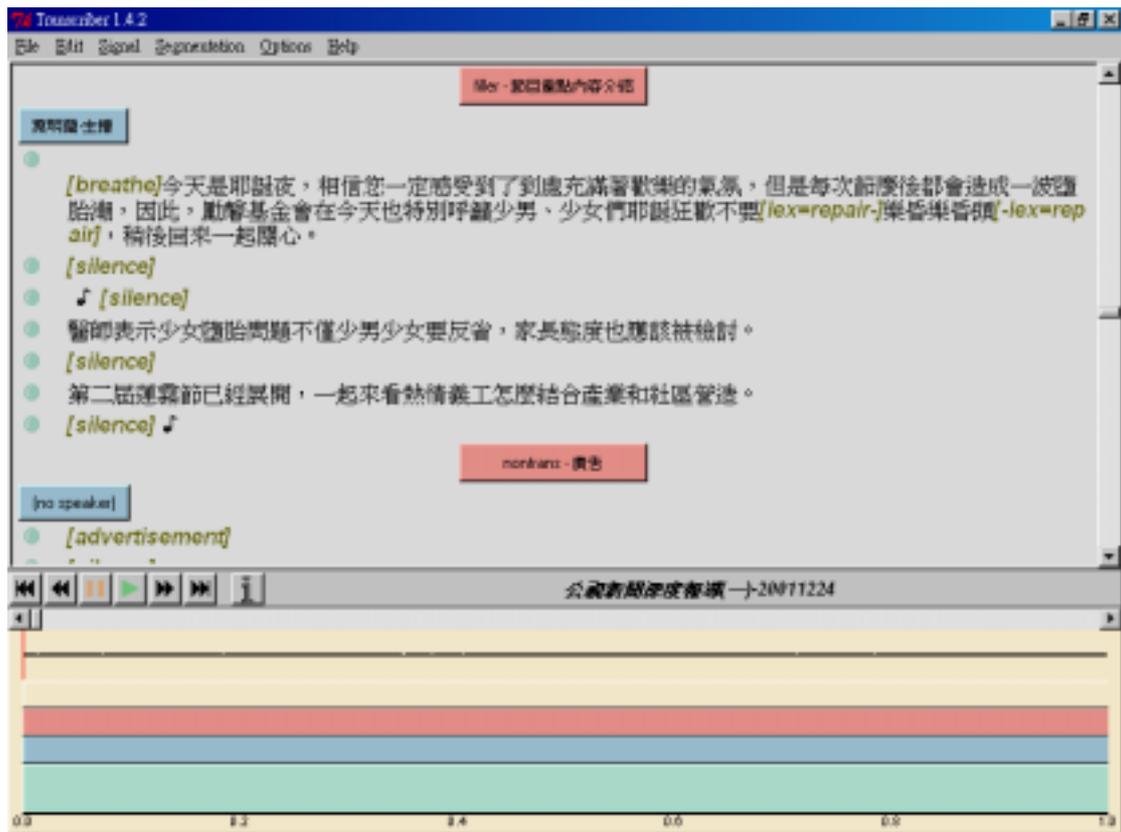
圖一：



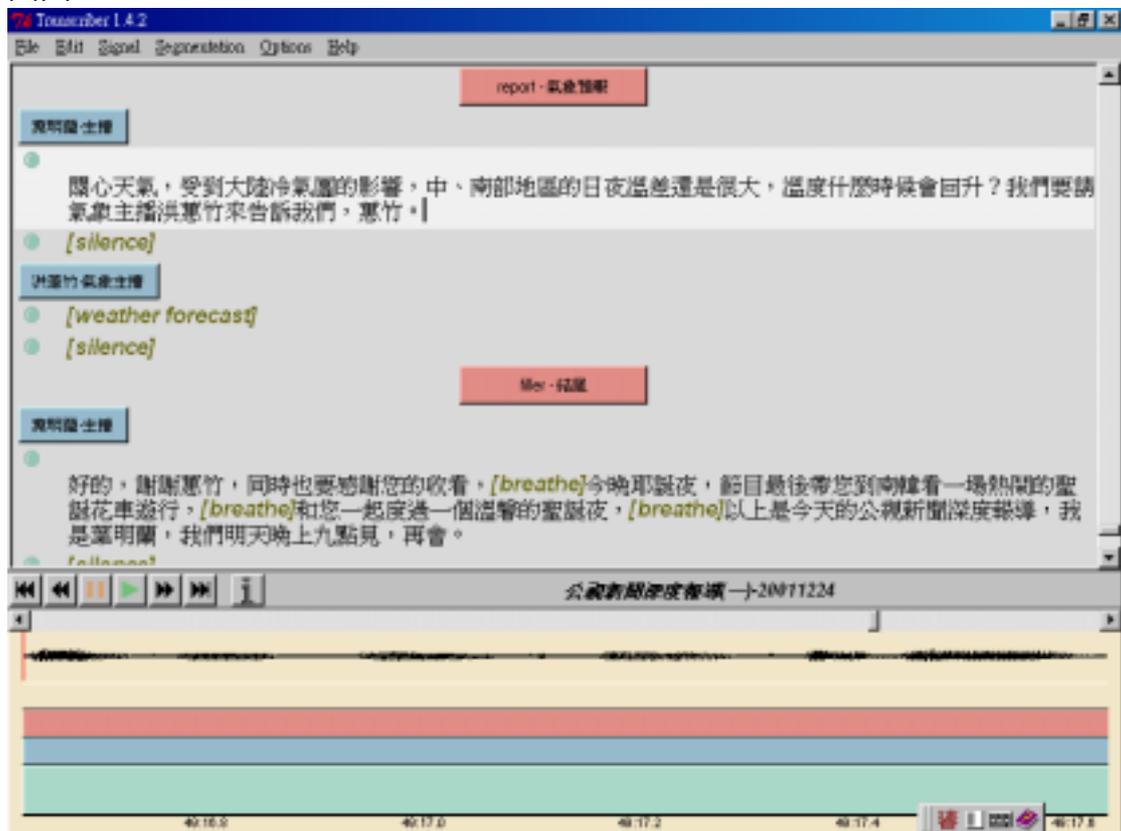
圖二：



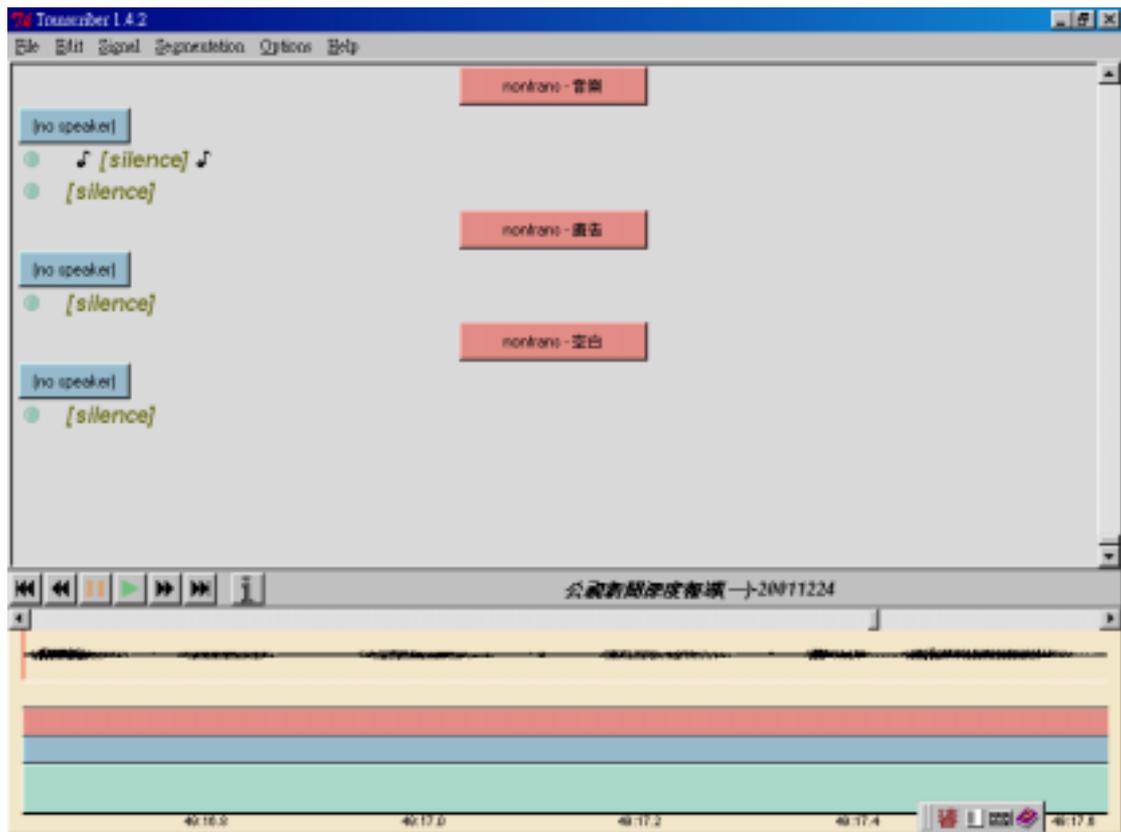
圖三：



圖四：



圖五：



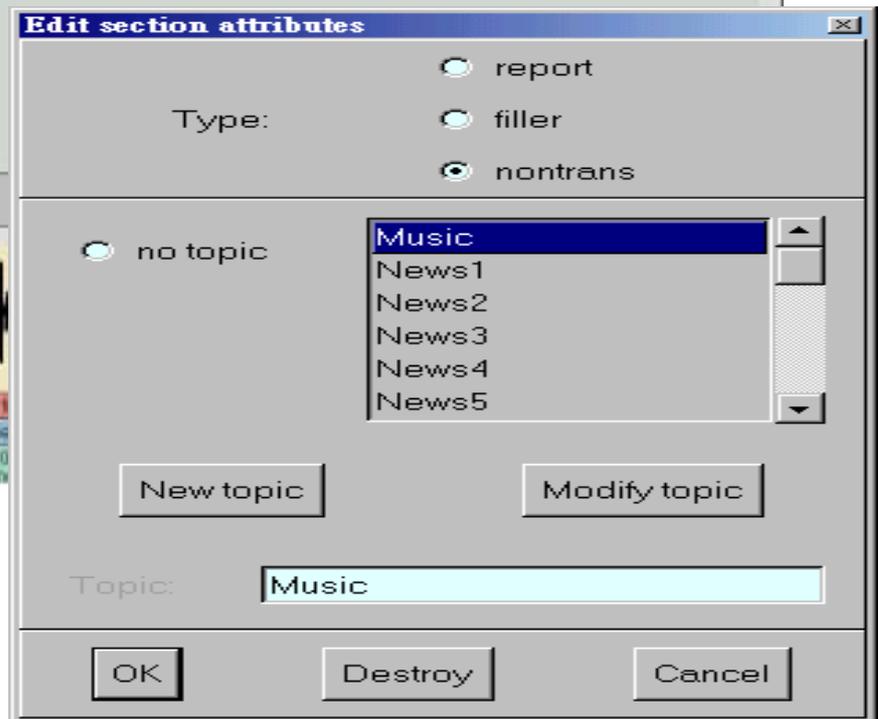
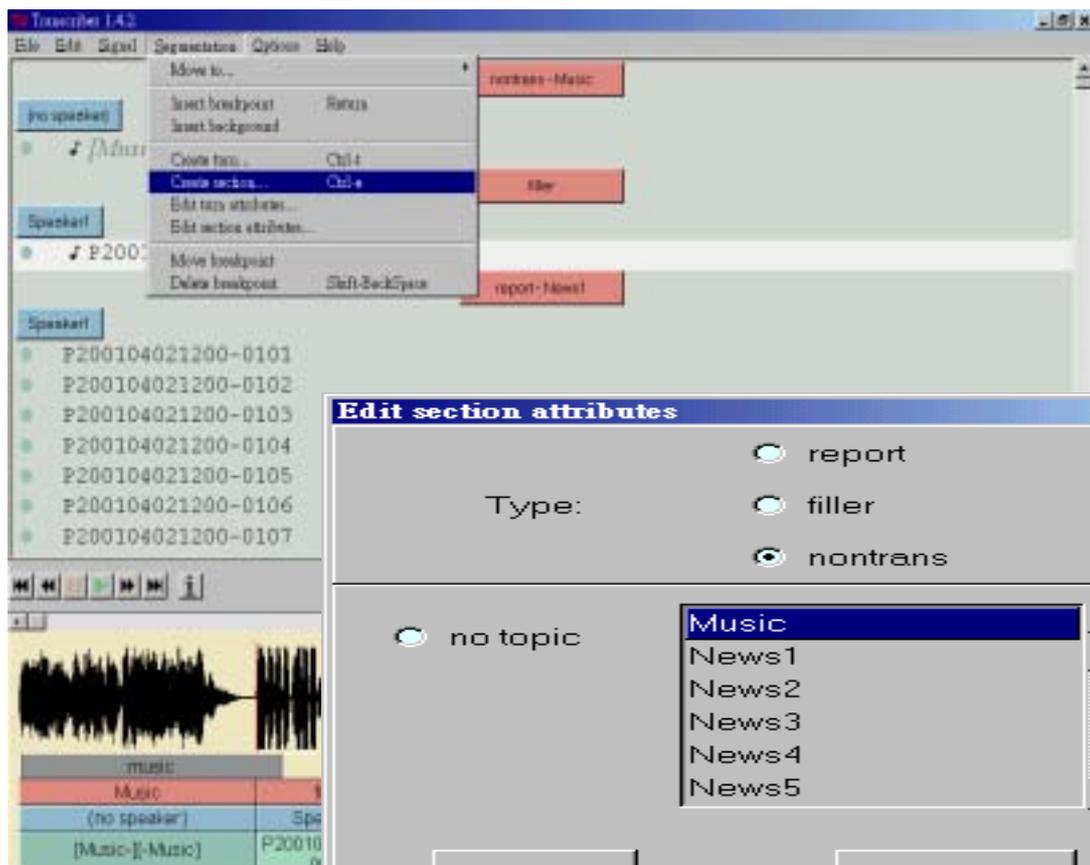
以上為『公視新聞深度報導』標註時各段落之基本架構。

3.2 語音資料標誌

一、段落主題

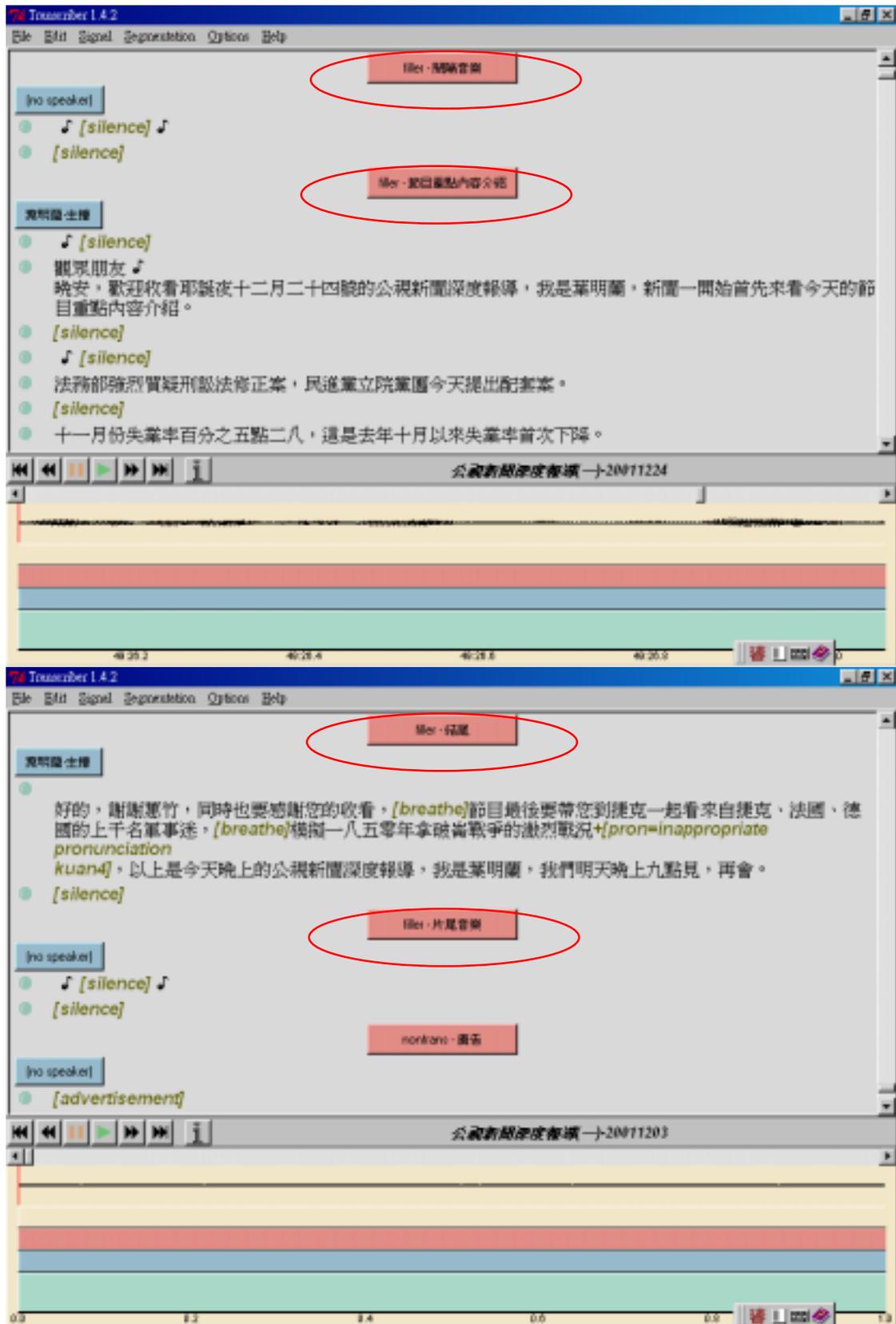
I. 段落主題標註方式如下：

點選 Segmentation 項目，選擇 Create section，在 Edit section attributes 視窗中建立段落種類與段落標題。

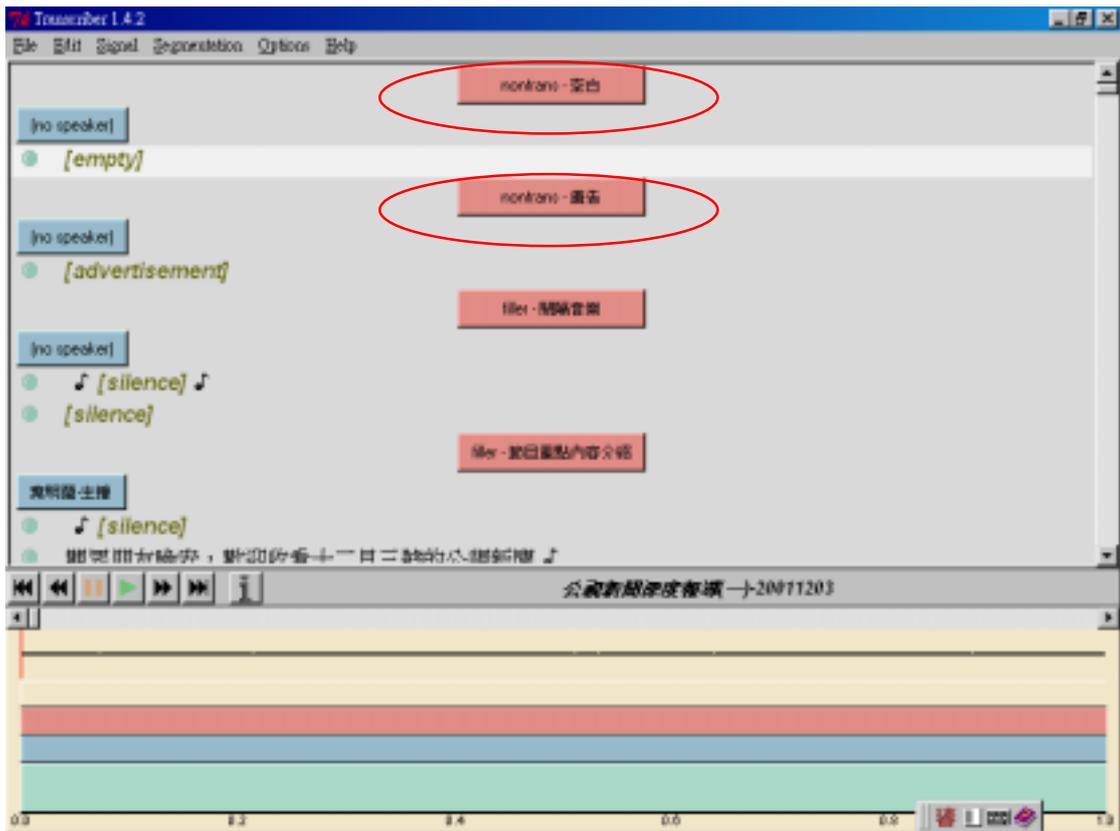




B. Filler :
 包含間隔音樂、節目重點內容介紹、結尾、片尾音樂等。



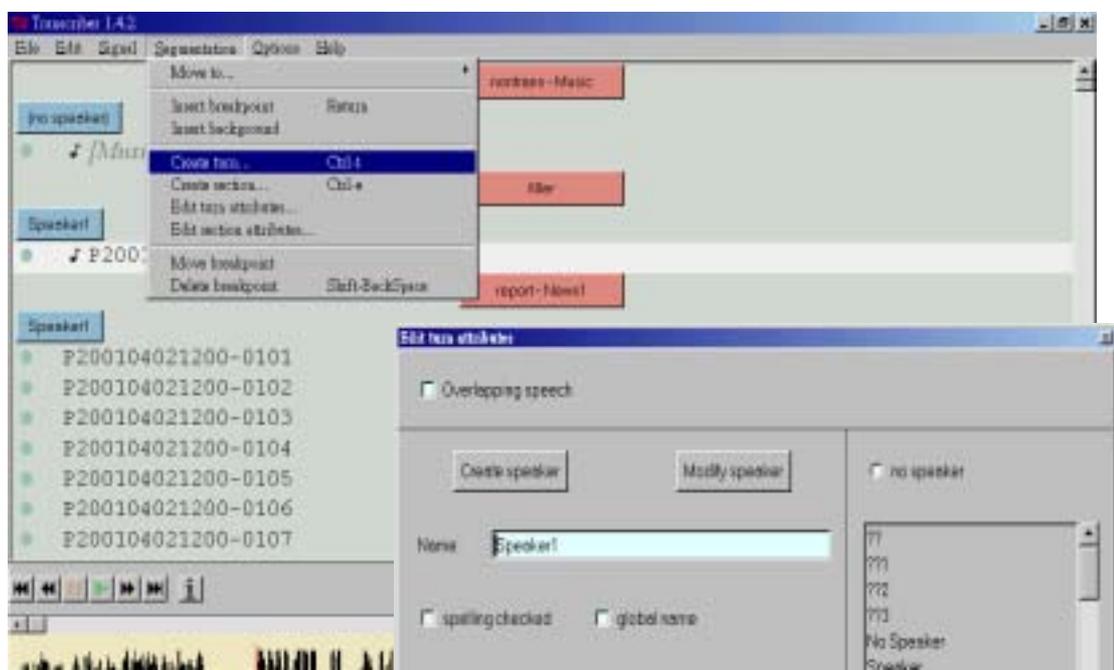
C. Nontrans :即不敘述的部分，包含空白、廣告。



二、說話者名稱

I. 說話者名稱標註方式如下：

點選 Segmentation 項目，選擇 Create turn，在 Edit turn attributes 視窗中建立說話者名稱及其相關基本資料。



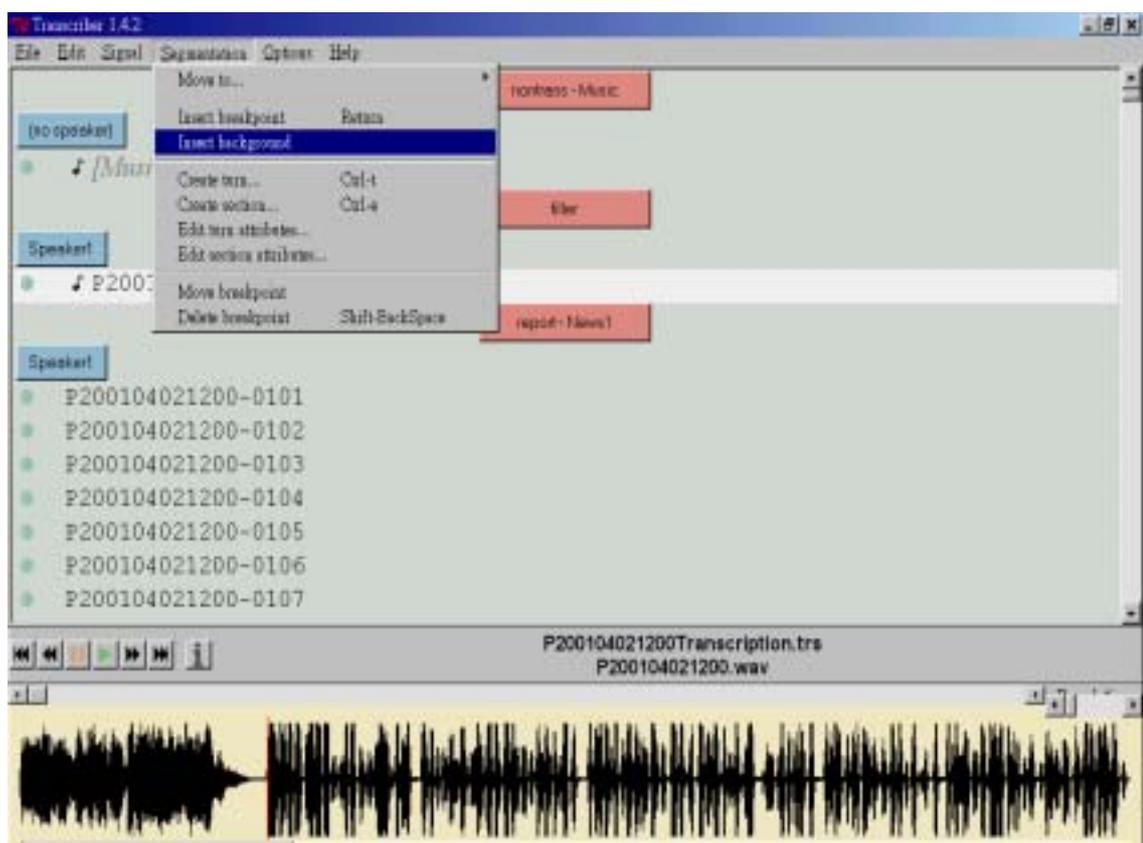
II. 標註重點

- A. 除主播及記者外，其餘說話者均只需標註姓名即可。
- B. 除主播、記者或可明顯聽出說話者為念講稿部分標註為 `planned` 之外，其餘說話者均標註為 `Spontaneous`。
- C. 空白、廣告部分之 `turn` 均標為 `no speaker`。
- D. 節目重點內容介紹部分不需更換說話者。
- E. 外籍人士一律設定為 `dialect="nonnative"`。

三、背景聲音

I. 背景聲音標註方式如下：

點選 Segmentation 項目，選擇 Insert background，在 background attributes 視窗中選擇背景聲音之種類。





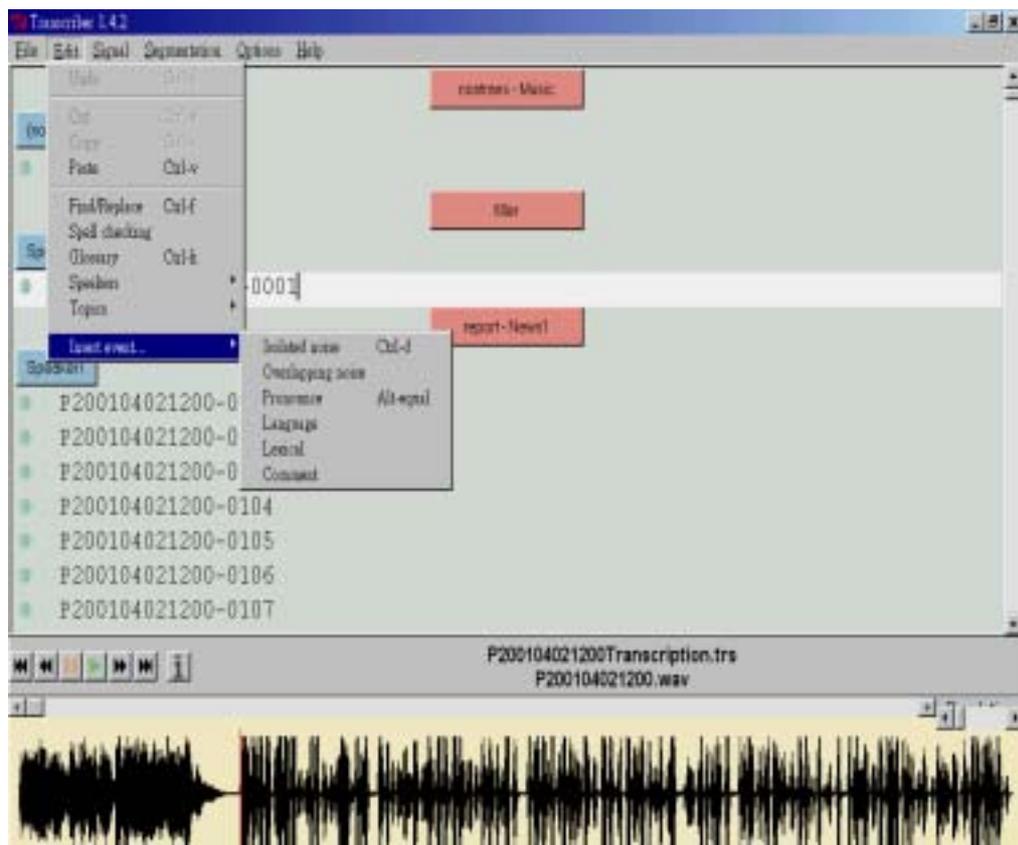
II. 標註重點

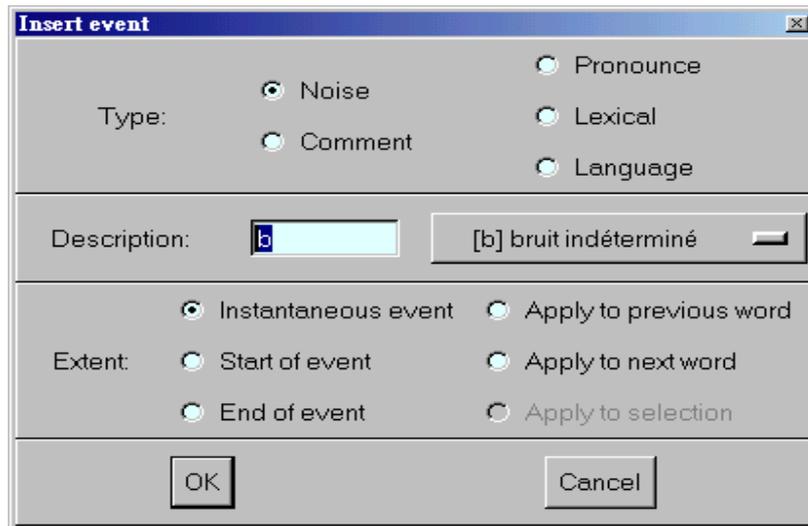
- A. 所有音樂均標為 background。
- B. background 之開始與結束的時間點盡可能與 breakpoint 相同。

四、 插入事件

I. 插入事件標註方式如下：

點選 Edit 項目，選擇 Insert event，在 Insert event 視窗中選擇事件的種類。





II. 標註之記號共五種類型，分別為”Noise”，”Pronounce”，”Lexical”，”Language”，”Comment”，其記號及標註方式說明如下：

A. Noise---

記號	說明	標註方式
advertisement	廣告	[advertisement]
breathe	喘息聲(含呼吸聲、吸氣聲、吐氣聲)	[breathe]
clear throat	清喉嚨聲	[clear throat]
click	嘖舌聲	[click]
cough	咳嗽聲	1 [cough-] 文字 [-cough] 2 [cough]
cry	哭聲	1 [cry-] 文字 [-cry] 2 [cry]
empty	DAT 轉錄至 PC 時,因無法同步作業而產生的 0 值 Samples	[empty]
hiccup	打嗝聲	[hiccup]
laugh	笑聲	1 [laugh-] 文字 [-laugh] 2 [laugh]
noise	非人所發出的語音或聲音	[noise]
particle	沒有標準語意的語氣詞	[particle-] 發音方式(大寫) [-Particle]

pause	停頓	[pause]
sign	嘆氣聲	[sign]
silence	沉默	[silence]
smack	咂嘴聲	[smack]
sneeze	噴嚏聲	[sneeze]
sniffle	吸鼻音	[sniffle]
swallow	吞口水聲	[swallow]
unrecognizable non-speech sound	由人發出非語音且無法辨識的聲音	[unrecognizable non-speech sound-]... [-unrecognizable non-speech sound]
weather forecast	氣象預報	[weather forecast]
yawn	哈欠聲	1. [yawn-] 文字 [-yawn] 2. [yawn]

B. Pronounce---

記號	說明	標註方式
alternative	尚未被收錄在辭典但被廣為使用之讀音	+ [pron=alternative 該字發音]
inappropriate pronunciation	發音雖有偏差但仍能辨識的字詞	+ [pron=inappropriate pronunciation 錯誤發音]
stutter	說話者口吃無法完整發音， 一直重複某個字或其部分的音	[stutter-] 重複的發音 [-stutter]
syllable contraction	說話者說得太快出現音節合併的現象， 不需標註發音方式	[syllable contraction-] 合併的字 [-syllable contraction]
uncertain	無法確定的字詞或含混音	[uncertain-]...[-uncertain]或 [uncertain-]該字發音[-uncertain]
unrecognizable speech sound	確屬人所發出之語音，但無法辨認何字 何意何音	[unrecognizable speech sound-]... [-unrecognizable speech sound]
zhuyin	音標	[zhuyin-] ㄅ ㄆ ㄇ [-zhuyin]

C. Lexical---

記號	說明	標註方式
abridged	中斷尚未完成的句子,重新開始一個新句	[lex=abridged-] 不完整的句子 [-lex=abridged]
cut	因剪接等技術關係導致說話者的言論中斷	[lex=cut-] 不完整的句子 [-lex=cut]
editing term	否定或放棄前面不完整句子的更正插語	[lex=editing term-] 更正插語 [-lex=editing term]
error	字詞,詞彙,成語或諺語錯誤使用,但不包括語音上的錯誤	[lex=error-] 錯誤詞語 [-lex=error]
interrupted	說話被另一方打斷,且說話權被搶走	[lex=interrupted-] 不完整的句子 [-lex=interrupted]
marker	說話者本身在語流中慣用的插入語,但此習慣插語已喪失其原有的意義	[lex=marker-] 插入語 [-lex=marker]
new word	新詞	[lex=new word-] 發音方式 [-lex=new word]
repair	說錯話後更正其內容	[lex=repair-] 錯誤及正確的詞語 [-lex=repair]
repetition	間斷性的重複詞語	[lex=repetition-] 所有完整重複的詞語 [-lex=repetition]
restart	因他人插話被打斷或因說話者自身的緣故重新開始一段詞語而有部分重複的現象	[lex=restart-] 所有部分重複的詞語 [-lex=restart] ex : [rstart-]陳水陳水扁[-restart]

D. Language---

記號	說明	標註方式
English	英語(單字需以英文表示,長句子則以...代表)	[lang=English-]。。。 [-lang=English]
Formosan	原住民語(內容用...代表)	lang=Formosan-]。。。 [-lang=Formosan]
Hakka	客語(內容用...代表)	[lang=Hakka-]。。。 [-lang=MHakka]
Japanese	日語(內容用...代表)	[lang=Japanese-]。。。 [-lang=Japanese]
Min-Nan	閩南語(內容用...代表)	[lang=Min-Nan-]。。。 [-lang=Min-Nan]
unknown	無法確定是何語言	[lang=unknown-]。。。 [-lang=unknown]

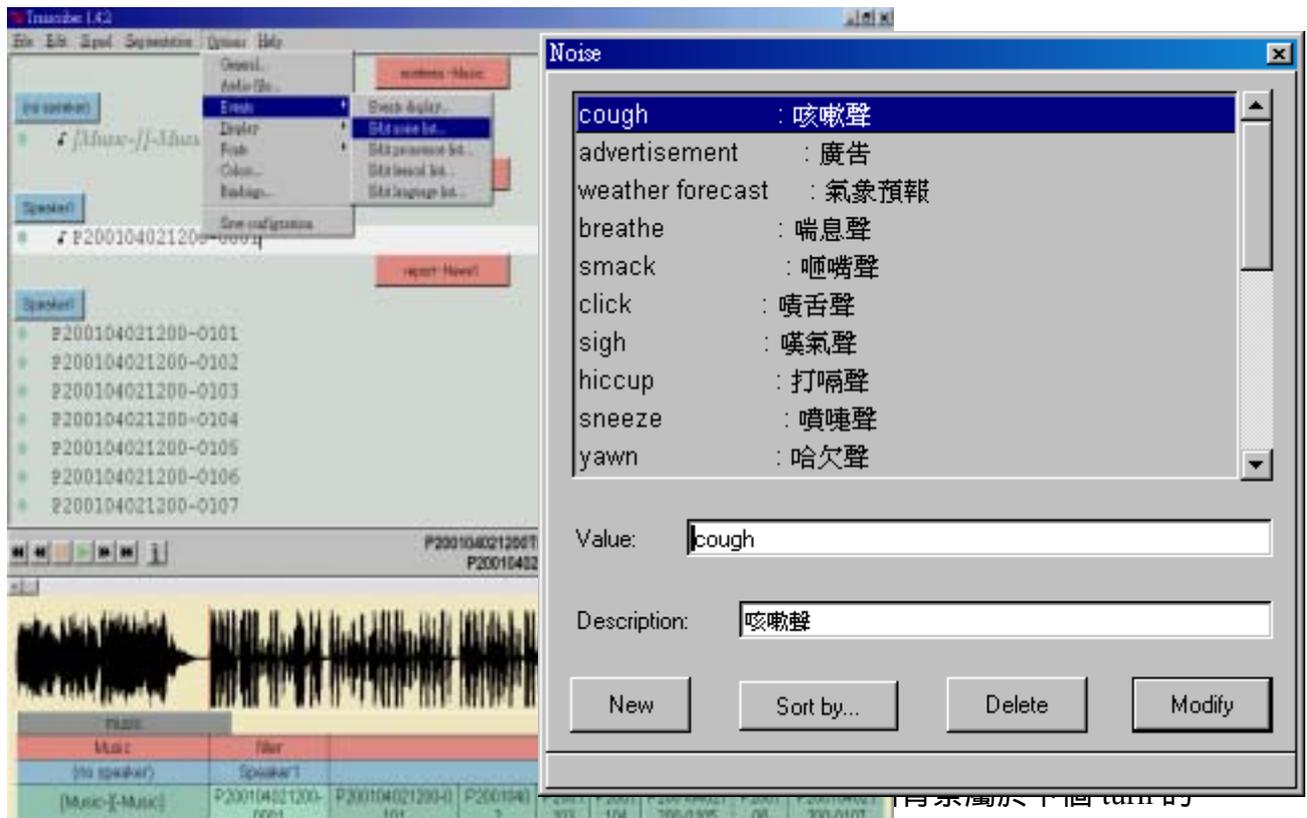
以上標註方法以中研院語言所曾淑娟老師的 Mandarin Conversational Dialogue Corpus 所製定的【口語標註規則與說明】為主,粗體字的部分是根據實際情況需要修改或增訂而成的。

E. Comment—一律標在 speaker turn 開始的最前端。 -

記號	說明	標註方式
low fidelity	聲音忽大忽小,在 turn/section 開始處插入 Comment,最好與 break point 重疊	{low fidelity from 28:45.474 to 28:46.858}
echo	出現回音時	{echo from 28:45.474 to 28:46.858}
the background sound is the preceding speaker's speech	適用的情況如下：攝影記者先進入立法院，錄下某立委的語音，然後記者才開始報導，此時立委仍在發言，此時某立委的語音變成背景聲音。	某立委 #xxxxxxx reporter # {the background sound is the preceding speaker's speech.} xxxxxxxxxxx
two kinds of background sound, second background sound started	適用的情況如下：有背景聲音的影片先出現後，然後採訪帶的內容才跟著出現。此時可能會有兩種背景聲音重疊(例如：先是大自然的蟲鳴，後來出現旁白，且旁白可聽出有明顯的背景雜音。)	reporter # silence # {two kinds of background sound, second background sound started.} xxxxxxxx(標註時以音量較大的 background 的類別為準。) #silence # {second (or first, 看是哪一個停掉) background sound stopped.} silence
The background sound is included due to the overlap of the speech and the following advertisement	適用情況如下：主播未說完廣告已經開始，標註方式是將廣告聲音視為 background，且 background 結束於廣告 section 開始處	{The background sound is included due to the overlap of the speech and the following advertisement.}
Min-nan-(Hakka-, Formosan-, Cantonese-, Other-dialect-, Foreign-language-)influenced pronunciation	受其他語言影響的不標準的聲音	Ex : {Min-nan influenced pronunciation from 28:45.474 to 28:46.858}
many speakers speak simultaneously	如有多人同時講台語	{many speakers speak simultaneously}

III. 編輯插入事件

點選 Option 項目，選擇 event，在各種類事件之視窗中編輯事件種類及說明。



部分，silence 一律歸在前一個 turn。

- C. pause：說話者自身語流中之停頓屬於 pause，說話者變換時，才使用 silence。
- D. inappropriate pronunciation：所有發音不標準均以此項標註，含台灣華語、客家華語及發音有偏差等。

第四章 實際操作問與答

A. 關於 speaker turn

1. native 及 non-native 是以中文為準：國人及記者在 Speaker attributes 處一律標成 native speaker，老外或華裔人士若聽的出來口音不似華語則一律標成 non-native
2. 同一 speaker 在一個 speaker turn 內如果 speaking mode 不同時標註方式如下：

spk#1 (mode="spontaneous") (依據同一 turn 內主要的 mode 標示)

#silence

#.....

#silence

#{mode="planned"} (此段落的 mode)

例如：記者讀稿(planned)之後。[silence]。問受訪者問題(spontaneous)。

缺點：mode 標註模糊。優點：不會增加 speaker turn。

3. 如果主播同時也是外場記者，屬於主播的部分 speaker 為 xxx-主播，外場記者的部分 speaker 為 xxx-記者。

缺點：speaker 的數目會增加。優點：將同一名記者在棚內與外場的語音區別開來。

4. 如果主播對前一則報導先補充說明再繼續下一則報導，因為兩段話分屬不同 stories，仍需要在 story boundary 處差入 speaker turn。例如：『為了更進一步關心 xxx，今天晚上 10 點我們的島節目會有深入的報導，請鎖定公視節目頻道。』但若為『關心完 xxx，接著我們來關心&&&』則不插入 speaker turn，只當作是下一則報導的開場白。

B. 關於 Background

標註的初期，一段 Background 必同時兼有 start time 和 end time，後來因為有無 Background 的標註標準從寬，發現有時 Background 出現不間斷的連續改變，至使前一段 Background 的 end time 即是後一段 Background 的 start time，但因為一個時間點無法同時標註兩種 Background(例如鏡頭轉回棚內主播的一瞬間)，以致於後期的語料中有一大部分的 Background 標註，在 Ultra edit 當中只出現一連串的 start time 而非兩兩一組的 start time 與 end time。

C. 關於 Overlapping

在標註的過程當中，不時會遇到有 overlap 的情形出現，依不同的情況，標註方法如下：

- a. 如果因為被 interrupted 而發生的 overlap 現象，一律從被 interrupted 的最前端開始以 overlapped speech 方式標註（維持完整語意段落）。

Spk#1+Spk#2

1.今天天氣很好，[interrupted-]我們出去[-interrupted]玩好不好？

2 好棒！

Spk#1 那我們走吧！

- b. 三個人同時說話時的標註方法，基本原則是：(a)盡量切成兩兩一起說話的段落，(b)無法切開的部分將主要語者標成 spk1，transcription 置於 1.xxxxx 其餘語者的 transcription 置於 2.

spk1+spk2+spk3

1. [unrecognizable-]...[-unrecognizable]

2. {the other two speakers speaks simultaneously} spk2:

[unrecognizable-]...[-unrecognizable]。 spk3:

[unrecognizable-]...[-unrecognizable]

- c. 具中文口譯的外語演講或事後由記者翻譯成中文之新聞播報其標註方法如下：

spk1+spk2

1.Chinese sentence

2.[lang=?-] ...[-lang=?]

spk1 為中文的部分，若為現場中文口譯則 mode 是 spontaneous，若事後由記者翻譯則 mode 是 planned。

- d. 遇到眾人同時回答問題時，依不同情形標註方式如下：

- (1) 異口同聲

spk1

#好不好玩？

某多名兒童

#{many speakers speak simultaneously}好玩！

- (2) 異口不同聲

spk1

#好不好玩？

某多名兒童

#{many speakers speak simultaneously}group1: 好玩！ group2:

不好玩！ Or#{many speakers speak simultaneously}spk1: 好

玩！ group1: 不好玩！

標成 spk 或 group 視其為單一語者或一群人決定。

- (3) 僅二或三人同時回答

多名語者指四人以上。若實際只有二或三人回答，依據原訂二或三人同時說話的方式標註。如『某兒童一 + 某兒童二 + 某兒童三』。

- e. 山谷回音標註方式：

- (1) non-overlapped

spk#1

#silence

#你好嗎？

#[echo-]你好嗎？[-echo] (noise 類)

(2) overlapped

spk#1+spk#1

#你好嗎？

#[echo-]你好嗎？[-echo] (noise 類)

f. 多人同時講台語的標註方式：

某多名成人

#{many speakers speak simultaneously}

[lang=Min-nan-]...[-lang=Min-nan].

D. 關於 Pronounce

1. 最常見的情形是因各種原因而發音不標準，如：四處林立的網路咖啡店要如+[pron=inappropriate pronunciation lu2]何生存。只要是發音不標準皆依前例標註，且不再另行註明發音不標準的原因。但如果是一長篇當中出現過多發音不標準，如受外省腔、方言或原住民語等影響，原則上還是需要一一標註其實際發音，若數量過多，則整段加上 comment 註明→{a lot of inappropriate pronunciations}。
2. 在標註的過程中，討論後有共識的少數特例會在該詞語後加註 Comment，例如：
 - (a) 落[pron=inappropriate pronunciation lau4]跑{口語化詞彙}
 - (b) 調漲五十[lang=Min-Nan-]趴[-lang=Min-Nan]{外來語：%}，調漲五十[lang=Min-Nan-]爬現[-lang=Min-Nan]{外來語：%}。
 - (c) 騎一台[lang=Min-Nan-]歐兜拜[-lang=Min-Nan]{外來語 autobike, motorcycle}。
 - (d) 一個[lang=Min-Nan-]歐里桑[-lang=Min-Nan]{歐里桑：外來語}。

E. 關於 Noise

1. Silence：[silence]原則上跟在前段後面，若很明顯屬於後段的開始，則標在後段開始處。
2. Breathe：[breathe]跟著後面的句子標註，不要自成一節。
3. Pause：
 - a. 段落（可能包含同一個人的很多句子）中語流中斷，一律標成 [pause]，不區分 pause 及 short break（原本存在於曾淑娟老師的【口語標註規則與說明】，但是因為 pause 及 short break 兩者之間不易區別，討論後決定，凡是語流間斷，皆以[pause]標註。）
 - b. 段落（可能包含同一個人的很多句子）中語流中斷，一律標成 [pause]，若更換語者，則標成[silence]。

(1)語流中斷隨即繼續，Pause 標註成：

..... , [pause]..... - 句子邊界

或

.....[pause]..... - 句中

** 若中斷時間有一個音節的長度或以上就要標註[pause]

(2)若語流終止後更換語者，則標註為 Silence 自成一。

Speaker1

#.....

#[silence]

Speaker2

#.....

4. Breath：[breath]前若有明顯的 pause 就標成：xxxx, [pause] [breath] xxx , [breath]前後若有明顯的 pause 就標成 xxxx, [pause] [breath] [pause] xxx
5. Particle：[particle]若不屬於 400 基本音節，則用聲母韻母組成新音節。又 restart or abridged 的尾巴如果發音不完整，則將該字（音節）標成 particle。
6. Story 之間如果有明顯的機器按鍵的切換聲音(如：照相時快門的聲音)，標註成[noise]，歸屬於前或後的 story 則視其是為結束前一則新聞或是開啟後一則新聞而定。標註方式如下：

Story 1

Spk 1

#.....

#[silence]

#[noise]

Story 2

Spk 2

#[silence]

#....

#[silence]

or

Story 1

Spk 1

#.....

#[silence]

Story 2

Spk 2

#[noise]

#[silence]

#....

#[silence]

7. Particle：語助詞或語氣詞。如：唉唷。統一標註成『唉唷』{其他可能寫法為 哎 }，目前啊、呀、呢、嗎、嘛、啦、吧、哇、喔、哦都標註成 particle(語助詞)，標註發音 (A, IA, NE, MA, MA, LA, BA, UA, O, O)。雖然實際上有些時候應該直接 transcribe 為中文字，例如當『疑問詞』或是『感嘆詞』時，但是在此一律將其視為 particle，標註發音。

F. 其他：

1. 對每一集錄音資料除了產生如：PTSND20020101.wav，TSND20020101.tras 外，另外再加上 PTSND20020101-log.txt，用以記錄該檔案修正的內容，格式如下：

PTSND20020101-log.txt

檔案名稱：

標註完成日期：

標註人：

第一次修訂日期：

檢查人：

修訂內容：

(1)時間點

原標註：

結論（新標註）：

(2)時間點

原標註：

結論（新標註）：

第二次修訂日期：(格式如上)

- 2.結尾之後的統一格式為：

filler-片尾音樂

δ [silence] δ

nontrans-廣告

[advertisement] (即使錄得太長包含下一個節目，仍一律歸到廣告)

nontrans-空白

[empty]

- 3.reporter id 有誤的標註方式：

(a)如果 reporter 與 video 顯示不符，若確知 reporter 是誰，依正確的 speaker 標，在 log.txt 記錄：

the reporter on the video: xxx

the correct reporter：xxx

(b)如果 reporter 與 video 顯示不符，但不確知 reporter 是誰，標成某女（男）

記者。在 log.txt 記錄：

the reporter on the video: xxx

the reporter unknown 或 the possible reporters : xxx , xxx

(c)若 video 上無資料且不確定 reporter 是誰，標成某女（男）記者，不需要在 log.txt 記錄，但若約略知道是誰，只是無法確定，在 log.txt 記錄：

the possible reporters : xxx , xxx

reporter 與 video 顯示不符一般會發生在採事後配音的報導。

4. 受訪者不知名，video 上亦無任何資訊，speaker turn 統一標註為：某老人一、某老人二、...、某成人一、某成人二、...、某學生一、某學生二、...、某兒童一、某兒童二、...。若有些許資訊則標註為：社區民眾一、社區民眾二、...。不知名的記者標註為：某記者一、某記者二、...。Merry: 2002/1/30 (含) 以後，Pido: 2002/6/5 (含) 以後即採用新標準。
5. 2002/07/01 開始更名為"晚間新聞"。新聞標題前的"類別標註"取消。
6. 移動 segment point 的方法：將游標移至要移動的切點，按下滑鼠中鍵（滾輪）即可拖曳至正確的時間點。
7. 在氣象預報之前，若主播再次提醒之前播報的新聞內容的注意事項，這個部分統一標註為"filler"，topic 則沿用原新聞 topic。若與前面的新聞無關，則另訂 topic。若在氣象預報之後則納入"結尾"。
8. 後期有一些新聞，會依照新聞的類別分 2-3 個大段落（如國內新聞及國際新聞），在各個段落的一開始都有一段相同的音樂，此段音樂目前標註在段落第一個 section 的開始位置，而該段音樂似乎應該比照『片尾音樂』自成一一個 filler section，但考慮有時候音樂未結束主播即已經開始播報新聞，自成一一個 section 可能發生與後面的 story section 重疊的現象，在沒有找到更好的解決方案之前，暫時維持原先的標註方式。
9. 自動標註軟體：

為了加速工作進度，設計了一自動標註軟體，可以將 section 大致的類別、標題及時間點自行標註完成。在使用自動標註軟體的過程中，發現了幾個常見的問題如下：

 - (1) 自動標註軟體輸出之.trs 檔案若 transcriber 在 loading 時有問題，只要將 transcriber 指出有問題的那一行最後面的空白移除即可。
 - (2) 部分自動標註檔案會有 section 的結束時間在後面 section 的開始時間之後的情形，遇此情形，只需至 Ultra Edit 將 section 的結束時間更改為後面 section 的開始時間即可。
 - (3) 有時自動標註軟體未對應到所有的新聞標題，導至標註者需上網尋找新聞稿或自行打字標註，造成作業上的不便，故後來決定將未對應到的新聞標題的內容先另存一個檔案(檔名.trs plus 檔名-unaligned.txt)。

附件二：電視節目錄音資料及對話語音資料處理技術報告

目錄：

1. 錄音資料來源
2. 使用軟體
 - 2.1 CoolEdit 2000
 - 2.1.1 CoolEdit 2000 功能簡介
 - 2.1.2 CoolEdit 基本使用說明
 - 2.2 Transcriber
 - 2.2.1 Transcriber XML 架構介紹
 - 2.2.2 XML 標籤介紹
 - 2.2.3 Type 詳細介紹
 - 2.2.4 Transcriber 使用說明
 - 2.3 UltraEdit
 - 2.3.1 UltraEdit 功能簡介:
3. 語音資料標示成果
 - 3.1 IC 電台標註狀況
 - 3.2 中央社標註狀況
 - 3.2.1 對話語料 2 中央社製播
 - 3.2.2 對話語料 3 中央社製播

1. 錄音資料來源:

語料來源包括：(1) FM97.5 IC 之音廣播電台，(2) 中央社製播語料庫。前者由廣播直接錄製，後者由中央社錄製。

2. 使用軟體：

語料之處理使用了三個軟體，包括：CoolEdit 2000、Transcriber 及 UltraEdit，各軟體之用途簡述如下：

- (1) 用 CoolEdit 分割所錄下來的新聞、廣播的 audio 檔成適當的長度，或是從中取出想要處理的部分。
- (2) 用 Transcriber 載入 audio 檔，編輯此 audio 檔的語音標註，存成*.trs 檔。
- (3) 可以使用 UltraEdit 或其他文書編輯軟體對*.trs 檔再做處理。

以下各節詳細說明各軟體。]

2.1 CoolEdit 2000 軟體介紹

CoolEdit 2000 程式是一個非常不錯的數位錄音程式，只要在電腦的音效卡上接一個麥克風就可以利用電腦來錄音，並且可以對此音檔做一些處理。

2.1.1 CoolEdit 2000 功能簡介

1. 數位聲音處理

CoolEdit 可以對錄進來的聲音做一些基本的處理，如：

財 **去頭去尾**：如果利用錄音機來錄音，錄音的過程中，開頭或結尾常常是按鍵操作的咖啦咖啦聲，或是無名的空白，這部份很容易用 CoolEdit 刪除。

財 **調整音量**：通常錄製完成後，音量的大小常常不是令人滿意，尤其是錄音效果差的設備，或是錄音時音源太遠而造成音量不佳的情況，可以利用 CoolEdit 很方便的調整音量。

財 **去背景雜音**：這是 CoolEdit 神奇而驚人的地方之一。背景雜音不外乎錄音設備錄音時所產生的高頻雜音(打開喇叭或是麥克風不是就可以聽到輕微的嘶嘶聲)，甚至錄音帶轉動都會有莫名其妙輕微的馬達聲、轉動聲，這部份的雜音可以利用 CoolEdit 幾乎完全去除。

2. 聲音的剪接

CoolEdit 的聲音檔案有非常精確的聲音長度標示，可以到達千分之一秒的準確度。一般的錄音機只有一個如同里程表的簡單三位數字標示，實在是非常的不精確。CoolEdit 還有一個好處是可以看到聲音音波的波形，從波形可以很容易的辨認出聲音、音樂的段落，甚至講了幾個詞都可以很容易

的辨識出來，要把一段話、甚至所講個幾個字選出來是很容易的。剪接的同時，CoolEdit 可以很容易的設定一些專業的剪接方式，如淡入(Fade In)、淡出(Fade Out)等等。

3. 製造特殊效果

- a.調整音調高低：女聲變男聲、男聲變女聲，如同現在正流行的 KTV 卡拉 OK 麥克風。另外唱 KTV 時，為什麼可以調整歌曲主調的高低也是這個道理。
- b.調整節奏快慢：將正常速度的聲音，變成機關槍、唐老鴨似的唧哩刮啦的快速講話的聲音；或者將正常速度的聲音，變成好像豬公他，慢慢...，緩...慢講話的聲調。
- c.回音：也可以很容易辦到。
- d.空間的音響效果：譬如同一個人在禮堂中講話和在小小的房間中講話聽起來感覺就是不一樣，這就是所謂的空間音響，也是回音所造成。在 CoolEdit 中，可以將原始的錄音很容易的處理轉換成在各種不同的空間中講話的效果，譬如在大教堂中或是空曠的草原上。

4. 混音、加背景音樂

利用 CoolEdit 可以很容易做「混音」的將兩種以上的聲音混合在一起，譬如將訪問講話的聲音，配上一段背景音樂，形成特殊的談話氣氛。或是混合其他的蟲鳴、鳥叫聲，讓人以為在戶外鄉野中採訪。

5. 處理各種聲音檔案格式的能力

CoolEdit 幾乎可以開啟、儲存各類常見的聲音格式檔案，可以算是一個非常好用的聲音格式轉換工具。新版的 CoolEdit 尤其可以將聲音檔案儲存成 MP3 格式。MP3 是目前最常見的聲音、音樂壓縮格式，可以將很大的聲音檔案，在不太損及聲音品質的情況下，將檔案壓縮成只有原來十分之一。

2.1.2 CoolEdit 基本使用說明

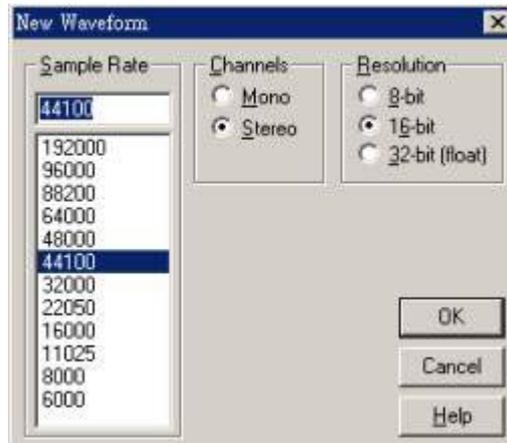
在畫面執行 CoolEdit 2000 圖示就可執行 CoolEdit 2000

進入畫面後，就可執行「錄音」、「放音」、編修「音樂」。



1、錄音(新建檔案)：

- A、按「File」、「New」或
- B、出現 New Waveform 畫面
- C、錄音鈕就可以用「麥克風」錄音。



播放：按播放鈕可播放音樂（聲音） 

連續播放：連續播放音樂 

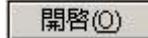
暫停：暫停音樂

停止鍵：停止播放 

2、開「音樂」檔：

- A、按「File」、「Open」或 
- B、出現「Open a Waveform」畫面，按「檔案類型」的 ▼，會出現各種「音樂」類型後，選定要播放的「類型」（如上例的 MP3）。



- C、選定「音樂」檔後，按「開啟」，開啟音樂檔 
- D、出現「讀入」過程後，畫面出現「音波」圖形。

播放：按播放鈕可播放音樂（聲音）  **停止鍵：**停止播放 

3、編修音樂 1（複製、貼上）：

- A、使用滑鼠點選要複製的起頭，拉動滑鼠到音樂檔的結尾，放開就是「選定」。（如下圖）



B、按「Edit」、「Copy」，完成「複製」。

C、可選「Edit」、「Paste to New」，複製成一個「新」音樂檔案。先用「Windows」，選定其他「音樂」檔，再選可選「Edit」、「Paste」貼在「音樂」檔之後。

同上如果選擇「Edit」、「Mix Paste」可作混音（合唱）功能。

4、編修音樂 2（刪除、取樣率修改）：

A、選定某一句話（或音樂），按「Edit」、「Delete Selection」將選定部分去除。

或選定某一句話（或音樂），按「Edit」、「Adjust Sample Rate」可以將取樣率作修整。

5、選擇「Transform」、「Invert」作高低音的相反改變，選擇「Transform」、「Reverse」作聲音由最後往前發聲。

6、選擇「Transform」、「Amplitude」後出現 Amplitude 功能表，選定「大、小」聲後，就可以將「聲量」變大或變小。

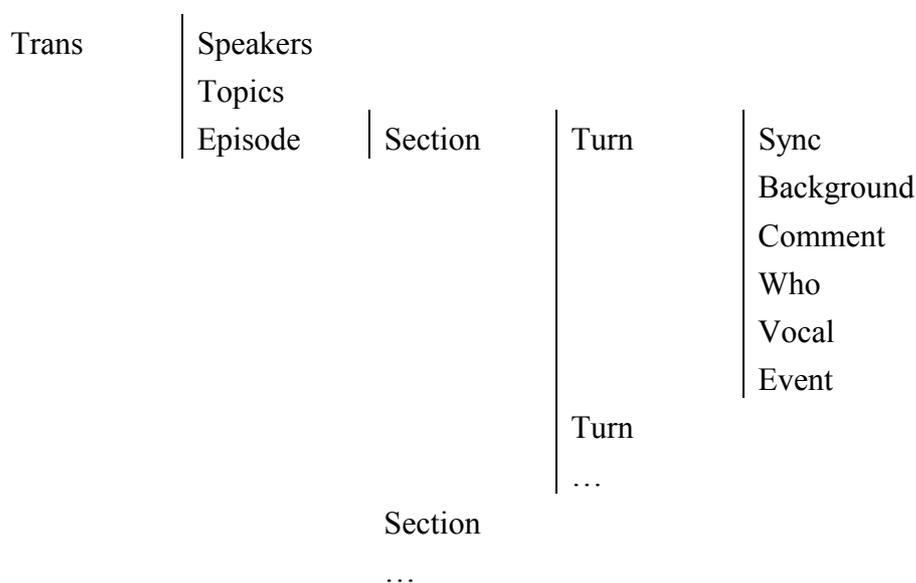
2.2 Transcriber 標記方法介紹

這是一套供 listener 將聽到的聲音紀錄下來的軟體，這個從聲音轉文字的步驟就叫做「Transcribe」。它可以紀錄時間、文字、特殊聲音及特殊事件等資訊。

2.2.1 Transcriber XML 架構介紹

以 Transcriber 產生的 XML 文件，其依照 XML 的定義為一個 tree 的架構（recursive structure，由 root node 與 sub tree 所組成）。

整篇 Transcriber 產生的 XML 語法如下所示：



2.2.2 XML 標籤介紹

1. Section (段落) 分成三類

- i. **Report**：新聞報導中主播與記者之播報內容及受訪者的訪談內容。
- ii. **Filler**：主播的引言或片頭、片尾音樂的部分。
- iii. **Nontrans**：即不敘述的部分，目前主要用再空白、廣告等部份。

2. Speaker 的各種標示：

- i. **Type**
 - 1 Male：男聲
 - 2 Female：女聲
 - 3 Unknown：無法藉由聽覺判斷男女的人聲或眾人齊聲說話。
- ii. **Dialect**
 - 1 Native：母語（中文）
 - 2 Nonnative：非母語（外籍人士均標示為 Nonnative）
- iii. **Mode**
 - 1 Spontaneous：自然流露的、不經計畫與強迫的
 - 2 Planned
- iv. **Fidelity**
 - 1 High
 - 2 Medium
 - 3 Low
- v. **Channel**
 - 1 Studio：錄音

3. Background 分成四類

- i. **Music** : 純音樂。
- ii. **Speech** : 可以聽清楚的人聲。
- iii. **Shh** : 機器聲。
- iv. **Other** : 噪音, 如交通工具的聲音 (呼嘯聲、喇叭聲、警鈴聲)、喧雜的人聲、電子通訊器材發出的干擾聲, 以及任何沒有意義的聲響。

Ps. 四種背景一次僅能選擇一種

4. Event 的種類

A. Type

- 1.Noise : 如打噴嚏、打哈欠、哭、笑、尖叫、吸氣聲或呼氣聲等。
- 2.Comment : 紀錄者的附註。
- 3.Pronounce : 發生的狀況, 例如發音不正確、低聲細語或縮寫等。
- 4.Lexical : 辭彙或語法使用發生錯誤或不清楚的時候。
- 5.Language : 出現非中文 (foreign) 的時候。

B. extent

- 1.Instantaneous event, 記號為[%s] : 短暫而瞬間發生的事件
- 2.Start of event, 記號為[%s-] : 與 end of event 搭配使用。及使用在事件發生的開頭。
- 3.End of event, 記號為[-%s] : 與 start of event 搭配使用。及使用在事件發生的結尾。
- 4.Apply to next word, 記號為+[%s] : 針對前一事件加以描述, 目前主要用在發音之部分。

C. Description

可以為任何資料, 如 marker、particle 等...

2.2.3 Type 詳細介紹

1. noise

- i. advertisement : 廣告
- ii. breathe : 喘息聲 (含呼吸聲、呼氣聲、吐氣聲)
- iii. clear throat : 清喉嚨聲
- iv. click : 漬舌聲
- v. cough : 咳嗽聲
- vi. cry : 哭聲
- vii. empty : DAT 轉錄製 PC 時, 因無法同時作業而產生的 0 值 Sample, 前後各約有數秒的時間
- viii. hiccup : 打嗝聲
- ix. laugh : 笑聲

- x. particle：沒有標準語意的語氣詞
- xi. pause：停頓
- xii. sign：嘆氣聲
- xiii. silence：沉默
- xiv. smack：砸嘴聲
- xv. sneeze：噴嚏聲
- xvi. snuffle：吸鼻音
- xvii. swallow：吞口水聲
- xviii. trill：顫音
- xix. unrecognizable non-speech sound：由人發出非語音且無法辨識的聲音
- xx. weather broadcast：氣象預報
- xxi. yawn：哈欠聲

2. Pronounce

- i. Inappropriate pronunciation：發音有偏差但仍能辨識的字詞（常見），判斷其拼音是否存在於漢語拼音中，若存在則使用新的拼音
- ii. Stutter：口吃，一直重複某個字或其部分的音，如「對對對」
- iii. Syllable contraction：說話太快而出現音節合併的現象（常見），如「這樣子」變成「降子」
- iv. Uncertain：無法確定的字詞，但是當一連串念了一句以後就可以辨別是什麼的字詞
- v. Alternative：尚未被收錄在辭典但被廣為使用之讀音
- vi. Unrecognizable speech sound：無法辨識的字詞，如方言
- vii. Zhuyin：注音符號（非常少用）

3. Lexical

- i. abridged：中斷尚未完成的句子，重新開始一個新句
- ii. cut：因剪接等技術關係導致說話者的言論中斷
- iii. editing：否定或放棄前面不完整句子的更正插語
- iv. error：字詞、詞彙、成語或諺語錯誤使用，但不包括語音上的錯誤
- v. interrupted：說話被另一方打斷，且說話權被搶走
- vi. marker：說話者本身在語流中慣用的插入語，但此習慣插語已喪失其原有的意義
- vii. new word：新詞
- viii. repair：說錯話後更正其內容
- ix. repetition：間段性的重複詞語
- x. restart：因他人插話被打斷或因說話者自身的緣故重新開始一段詞語而有部份重複的現象

4. Language

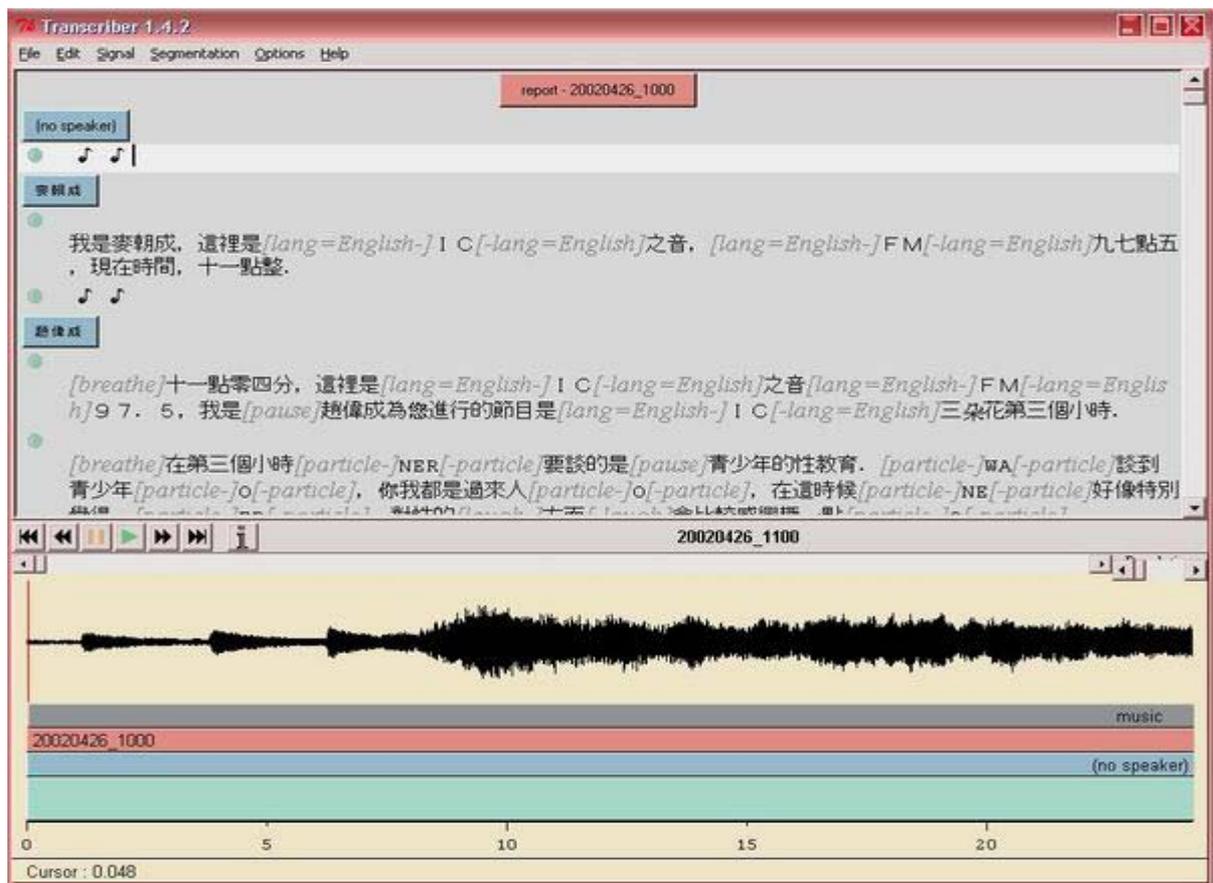
- i. Cantonese：廣東話（用...代表）
- ii. English：英語（單字須以英文表示，長句則以...代表）
- iii. Formosan：原住民語（用...代表）

- iv. Hakka：客家語（用...代表）
- v. Japanese：日語（用...代表）
- vi. Min-Nan 閩南語（用...代表）

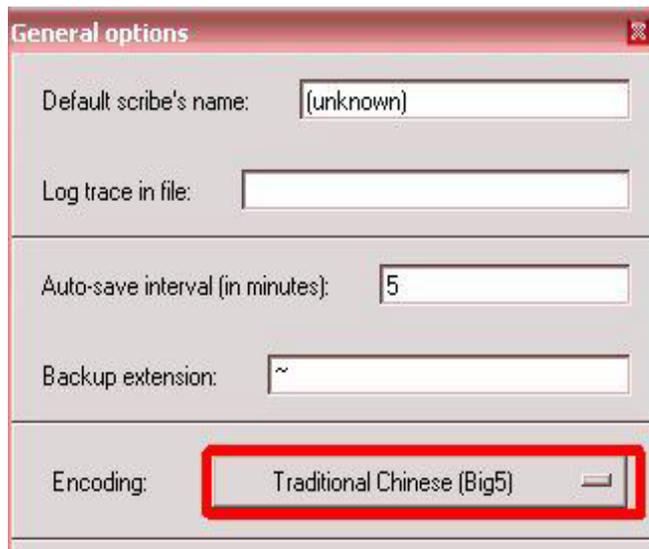
5. Comment

- i. Low fidelity from **:**:** to **:**:**：聲音忽大忽小，在 turn/section 開始處插入 Comment，最好語 break point 重疊

2.2.4 Transcriber 使用說明



- (1) 設定環境：在 Options\General options 裡的 Encoding 選擇 Traditional Chinese (Big5)，才能正確地用繁體中文編寫標註內容。



(2) 在 File\Open audio file... , 選擇載入所要處理的音檔。

(3) 在 File\Save as... , 把編輯好的標註內容存成*.trs 檔。

(4) 如何標註：

A . 先對各種代號下定義 (可作可不作)：

標註記號共分成四種插入事件，每種各有十多種不同意義的代號，可以自己先在 Transcriber 中加入。在 Options\Events\裡選擇編輯 noise , pronounce , lexical , language 四種其中一種：



先按「New」增加新的欄位，Value 是所要用的代號，Description 是此代號的意義。最後按「Modify」確定。

B . 開啟一段新的 audio 檔後，標註內容是一片空白。

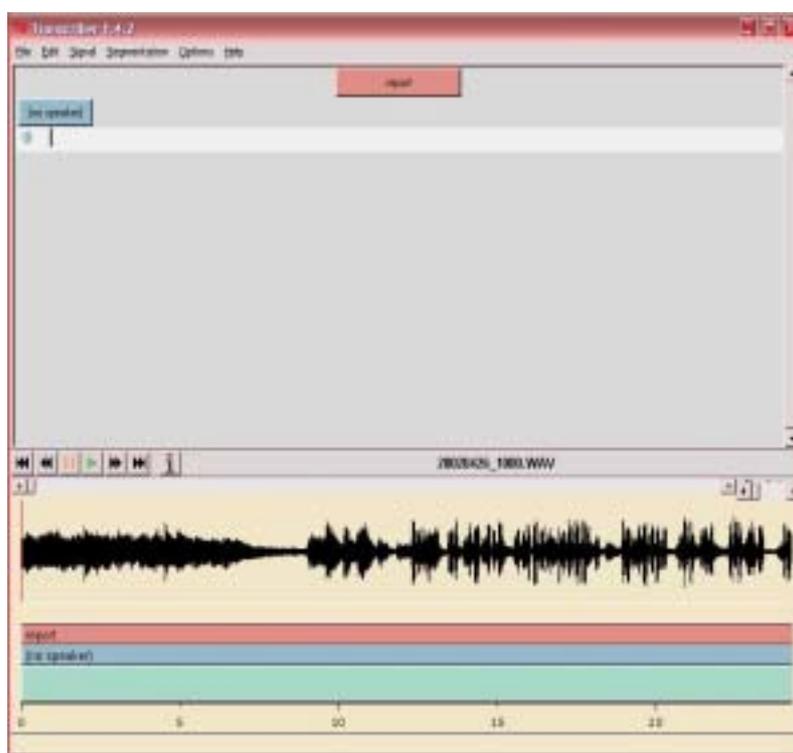
選擇標註內容的原則是：

只標註主持人與來賓（交談者）的內容。

準點報時，有標註。

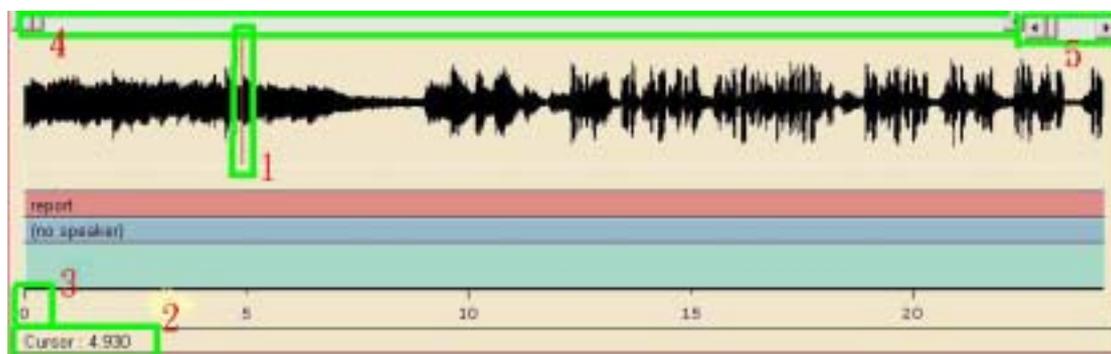
廣告，即時新聞，不標註。

音樂（音樂性節目或是串場用），不標註。



C . 基本操作：

下半部是 audio 檔的波形內容，皆以滑鼠操作。

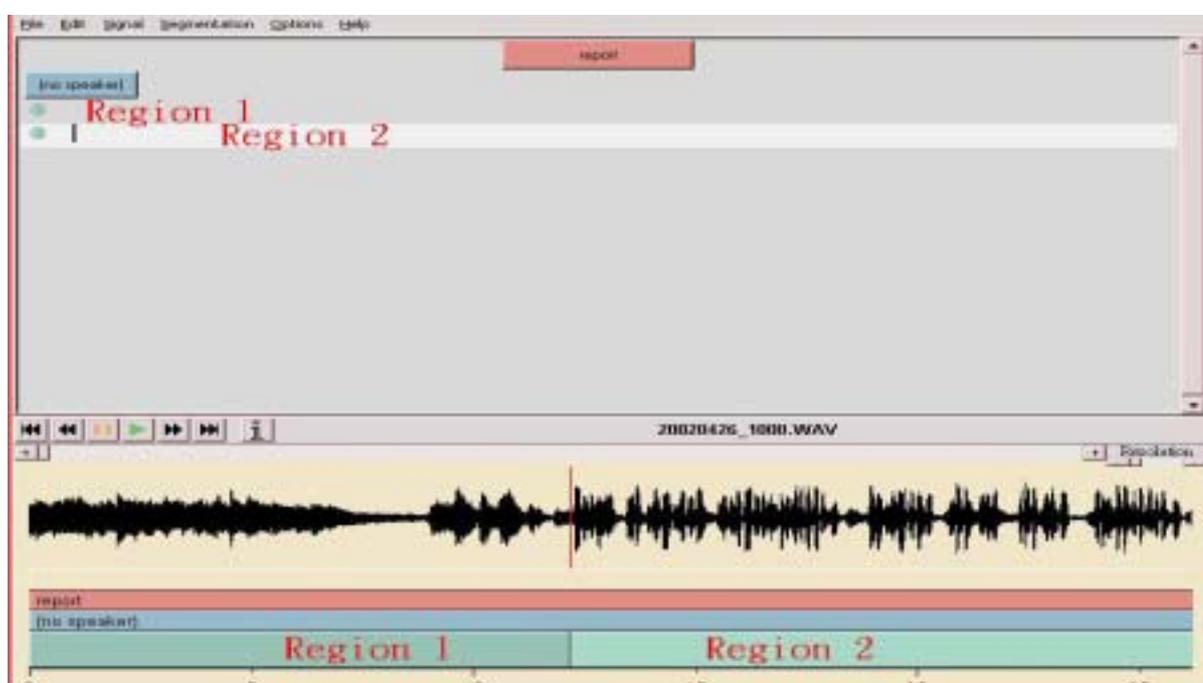


用滑鼠左鍵可選擇波形任何一點，會出現 1 的紅線，在 2 會出現對應的時間點。3 是目前在整段 audio 檔的實際位置，單位是秒。4 可以左右拉動看接下來的波形。5 是調整顯示的比例。

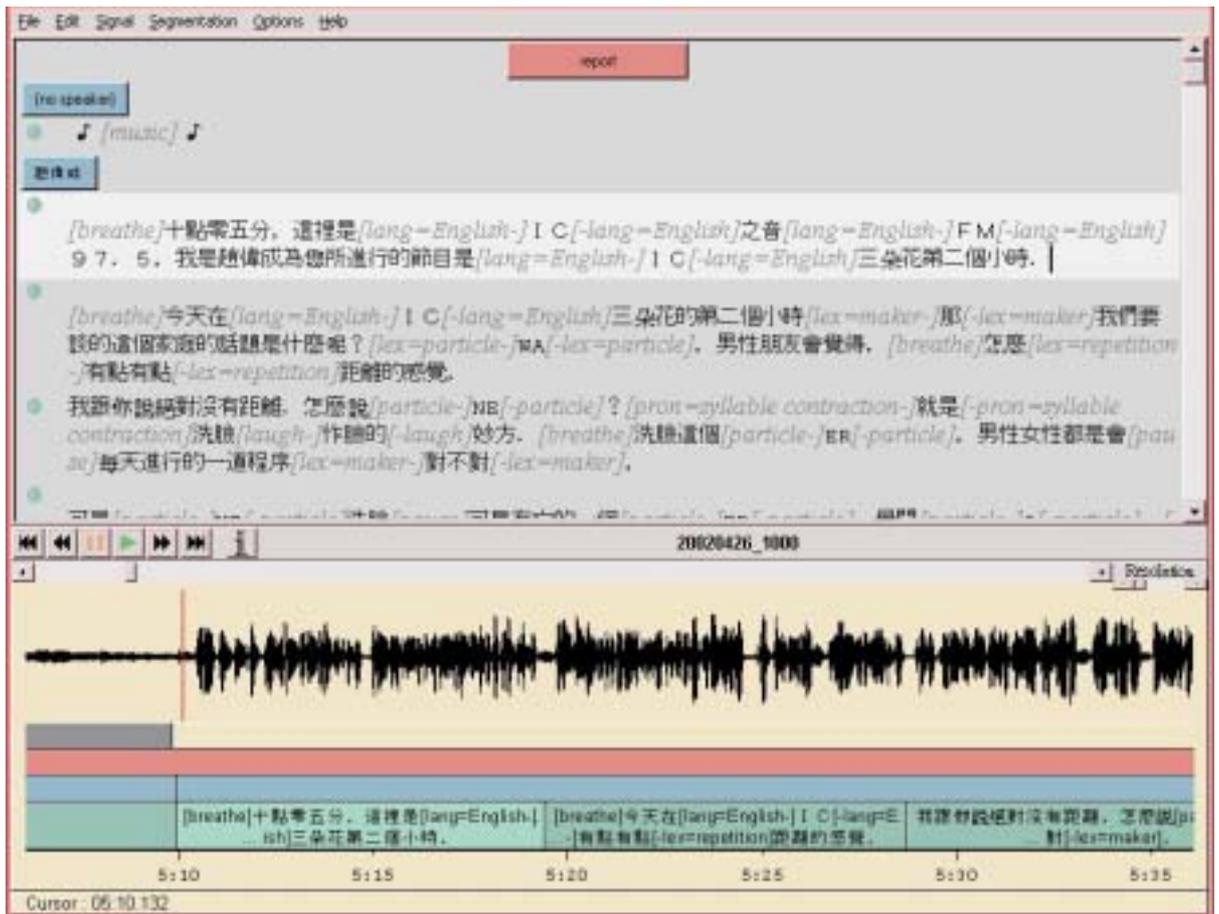
D . 如何標註：

所謂標註是先紀錄 audio 檔的文字內容，再根據說話者的語氣變化，在字與字之間加入說明的代號，稱作「插入事件」。

- (a) 首先在下半部波形的一段對話前，用滑鼠左鍵按出「起始點」，再按一下「Enter」作出一個 break point。由下圖可以看出，左半部是 music，右半部是語音，按一下「Enter」後，在文字工作欄位多出了 Region 2。同樣地，也要在一段句子的結尾也作出一個 break point，而文字就寫在這一段 Region 裡。



例：下圖可以看到有 4 個區域，音樂不作標示，文字的部分，依照適當的長度再作分割。



(b) 通常處理的流程是：

- ◆ 先把整段音檔其中要標註（語音）與不要標註（音樂、廣告．．．）用 break point 大致分割出來。
- ◆ 聆聽每段語音的部分，把內容紀錄下來。
- ◆ 在聽一次語音，這次把各種插入事件標註進去。
- ◆ 依照長度規則，把語音每隔一小段時間再斷一次。
- ◆ 編輯細部部分，包括說話者名稱，背景音樂的標示。

此流程並不是標準程序，可以依喜好調整。

E . 最後的音檔標註文件應該包括：

- ◆ 文字內容以及各種插入事件的標註。
- ◆ 談話者的資料編輯（Segmention\Create turn）。
- ◆ Topic 的編輯，裡面有日期、時間、節目。

F . 各種工作準則：

有幾樣工作的原則以及我們自訂的一些規範：

a . 標註內容的選擇：

- ◆ 只標註主持人與來賓（交談者）的內容。

- ◆ 準點報時，有標註。
- ◆ 廣告，即時新聞，不標註。
- ◆ 音樂（音樂性節目或是串場用），不標註。

b. 插入事件的代號：

一律以中研院所給我們的 ” Transcriber 操作暨標註方式說明 ” 為準，沒有再增加或修改定義。

c. 語氣詞須以英文字母代替：

由於中研院的文件裡並沒有完備的對照表，所以我們自行訂出一組常見語助詞的對照表：

啊	吧	啦	呢	喔	嗎、嘛	嗯	呃、兒	呵	哈
A	BA	LA	NER	O	MA	N	ER	HR	HA

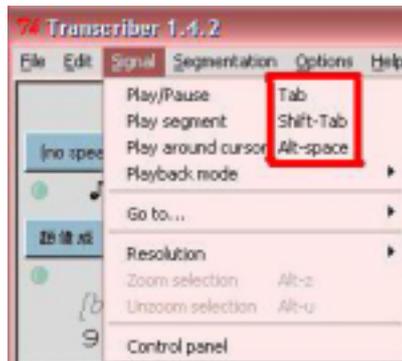
哇	ㄟ、ㄝ	囉	唉呀	唉唷	ㄗ ㄛ	ㄛ ㄛ	厚	呀	耶
WA	E	LO	AI YA	AI YO	M	NHM	HO	YA	YE

d. 切斷句子的長度：

為了作語音辨識的方便，把最後切成的每一小段句子的長度限制在 10 到 15 秒之間，最好是一個完整的句子，切斷點是在波形間隔的部分。句子的開頭與結尾處也最好能保留一些沒有語音的部分，不要剛好切在剛開始發聲的部分。（通常出現在一大段話的開頭與結尾部分，以及談話中有停頓或休息的地方）

G. 建議事項

- a. 善用快速鍵。一些常用的鍵像「Play」、「插入 noise」等可以用快速鍵控制，可以加快工作速度。可惜不能自訂快速鍵。快速鍵的看法：



H. 參考：

1. 中研院 ” Transcriber 操作暨標註方式說明 ”。
 - (1) 第一部分介紹指令欄內各項指令的意義，以及某些指令的設定方法。
 - (2) 第二部分是所有插入事件的介紹，包括所使用的記號、意義說明、標記方式及使用範例。
 - (3) 第三部分是簡單的實際範例。

2.3 UltraEdit 軟體介紹

除了可以像 Word 一樣，提供強大的編輯與列印功能外，如果用它開啟非文字檔，UltraEdit 就立刻成為可以直接編輯程式碼的編輯器。而且 UltraEdit 也支援雙位元字元集 (DBCS)，可編輯中文。

2.3.1 UltraEdit 功能簡介：

- (1) 提供超強的文字檔編輯、預視、列印功能。
- (2) 提供直接編輯十六進位碼功能。
- (3) 可編輯 HTML 檔案，以彩色顯示 HTML 標記，方便網頁編輯。
- (4) 新版在畫面左邊提供快速檢視欄框，無論切換編輯視窗或檔案都很容易。
- (5) 新增的 project (計畫) 功能，可以讓您把數個文字檔合成一個「計畫」，方便您一次編輯數個彼此有關連的文件。

3. 語音資料標示成果

3.1 IC 電台標註狀況

- (1) 以下的檔案是未將廣告、音樂...等非語音去除的部份

編號	節目名稱	檔名	總長度	中文音節數	Speaker 數
1	IC 三朵花	20020426_1000	59 分 4 秒	3044	2
2	IC 三朵花	20020426_1100	55 分 2 秒	5903	4

3	IC 晚報	20020425_1800	59 分 36 秒	10540	10
4	IC 晚報	20020425_1900	60 分 10 秒	10060	5
5	歡喜一家親	20030405_1	59 分 13 秒	7081	3
6	歡喜一家親	20030405_2	59 分 3 秒	5053	3
7	愛上新竹	20030405_3	60 分 4 秒	9887	3
8	愛上新竹	20030412_3	52 分 50 秒	9015	3

(2) 以下是已去除語音並將節目切成多個小段的部份

編號	節目名稱	檔 名	總長度	中文字數	Speaker 數
9	歡喜一家親	20030412_1_1	1 分 41 秒	452	2
		20030412_1_2	3 分 27 秒	494	4
		20030412_1_3	3 分 54 秒	474	3
		20030412_1_4	4 分 02 秒	977	3
		20030412_1_5	3 分 42 秒	1013	3
		20030412_1_6	3 分 44 秒	879	3
		20030412_1_7	1 分 11 秒	288	2
10	歡喜一家親	20030412_2_1	5 分 49 秒	1184	3
		20030412_2_2	5 分	1201	3
		20030412_2_6	50 秒	226	2
		20030412_2_7	1 分 34 秒	235	2
		20030412_2_8	1 分	191	2
11	心靈故鄉	20030804_1_1	6 分 51 秒	1421	2
		20030804_1_2	8 分 25 秒	1908	2

說明：

- A. 以上皆為已標註完成之音檔，編號代表一個一小時節目。
- B. 檔名命名方法，如：20030412_1_5
20030412 為錄音時間，(2003 年 4 月 12 日)
_1 為當天錄的第一個節目(每個節目一個小時)

_5 為節目切割後的第 5 個音檔 (切割目的為將非語音部分去除), 若無則表示音檔沒有切割過

C. 未完整標註完的節目 --- 編號 10, 11

編號 10 少了 20030412_2_3、20030412_2_4、20030412_2_5

編號 11 少了 20030804_1_3、20030804_1_4

D. 文字標註所使用的軟體：transcriber 1.4.2

(3) 節目性質說明：

A. 歡喜一家親、愛上新竹、心靈故鄉

皆是訪談性節目，主持人及來賓對於某一主題做深入討論

B. IC 晚報

前半播報新聞，後半有主播和記者...等對於一些新聞的討論

C. IC 三朵花

由兩三個人輪流出來談一些生活常識

3.2 中央社標註狀況

3.2.1 對話語料 2 中央社製播

編號	節目名稱/檔名	總長度	音節數	Speaker 數
1	居家 2	10 分 10 秒	2965	2
2	居家 3	9 分 55 秒	2911	2
3	居家 4	10 分 58 秒	3145	2
4	新聞	10 分 36 秒	3271	2
5	溫泉	35 分 49 秒	9654	2
6	運動	10 分 00 秒	2871	2
7	算命 5	10 分 15 秒	2512	2
8	算命 6	10 分 28 秒	2975	2
9	談狗狗	10 分 37 秒	2598	2
10	旅行與音樂 1	17 分 16 秒	3091	2
11	旅行與音樂 2	15 分 41 秒	3123	2
12	旅行與音樂 3	15 分 08 秒	2332	2
13	旅行與音樂 4	15 分 09 秒	2736	2
14	旅遊 1	10 分 13 秒	3115	2
15	遊遊 2	11 分 15 秒	3267	2

總數 15 個 總時間 203:30

節目性質與標註狀況說明

(1) 居家 2 :

一男一女的對話，很開心的語調在討論租房子的問題。基本上這兩位錄音者講話都非常清楚，也沒有很多無異議字。

種類	breathe	clear throat	laugh	silence	pause
個數	15	1	22	1	37
百分比	19.7	1.3	29	1.3	48.7

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 306)個數	59	77	13	24	5	59	11	4
百分比	19.3	25.2	4.3	7.8	1.6	19.3	3.6	1.3
	particle 種類	嗎、嘛	吧	啦	其他			
	個數	16	9	16	13			
	百分比	5.2	2.9	5.2	4.3			

由於男女兩方高興的談論租房子，所以在對話中笑聲的比率相對的會比較高。

(2) 居家 3

一男一女的對話，主要都用抱怨的語調在討論台灣房子間隔太小和租房子的一些事情。這段對話男女生重複的地方並不算多，可清楚的標註。

種類	breathe	click	cough	laugh	sign	silence
個數	59	1	4	22	2	3
百分比	48.4	0.8	3.3	18	1.6	2.5
種類	smack	swallow	pause			
個數	2	7	22			
百分比	1.6	5.7	18			

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 321)個數	54	64	28	30	4	44	6	7
百分比	16.8	19.9	8.7	9.4	1.3	13.7	1.9	2.2
particle 種類	耶	嗎、嘛	吧	啦	厚	其他		
個數	16	30	10	8	9	11		
百分比	5	9.4	3.1	2.5	2.8	3.4		

很明顯的，抱怨的語調中，常常會有嘆氣，所以 breathe 的次數會比其他多。

(3) 居家四：

一男一女的對話。剛開始主要是女生在**抱怨**的說買房子時會遇到的種種。後來換男的在討論買家電的經驗。基本上他們**說話清楚**，只是笑聲有時候會蓋過另外一個人的聲音，造成難以辨識說些什麼。

種類	breathe	clear throat	laugh	silence	smack	sniffle
個數	47	4	16	3	1	1
百分比	31.5	2.7	10.7	2	0.7	0.7
種類	swallow	pause				
個數	1	76				
百分比	0.7	51				

particle 種類	喔	啊	ㄟ、世	ㄊ 嗯	嗯	哇	嗯哼	呃、兒
(共 348)個數	67	60	24	2	39	3	82	12
百分比	19.2	17.2	6.9	0.6	11.2	0.9	23.6	3.5
particle 種類	呢	嗎、嘛	吧	啦	厚	囉	其他	
個數	7	29	5	10	4	1	3	
百分比	2	8.3	1.4	2.9	1.2	0.3	0.9	

跟上面一樣，大多都是 breathe 與笑聲。

(4) 新聞：

一男一女的對話。他們主要是用**批評的語調**在討論新聞的不專業。音檔中有一些英文的簡稱，像是 SNG，必須知道它是代表什麼意思才能前後連貫。有時候剛笑完要講下一句話時，通常會伴隨著剛剛的笑聲，聽的會很模糊。

種類	breathe	clear throat	laugh	pause	smack	sniffle	unrecognizable
個數	55	4	37	152	4	1	1
百分比	21.7	1.6	14.6	59.8	1.6	0.4	0.4

particle 種類	喔	啊	ㄟ、世	ㄊ 嗯	哇	嗯哼	呃、兒	呢
(共 223)個數	23	54	6	52	1	21	23	9
百分比	10.3	24.2	2.7	23.3	0.5	9.4	10.3	4
particle 種類	嗎、嘛	吧	啦	囉	其他			

個數	17	1	3	1	12
百分比	7.6	0.5	1.4	0.5	5.4

此對話出現的 pause 相當嚴重，遠比其他雜音多上三倍。

(5) 溫泉:

一男一女的對話。男生一開始沒有很專心的談話(在玩MSN)。後來兩人開始認真的談話。其中兩人的對話都是以開心的語調去對談，談論有關於一些家人的八卦和彼此最近發生的事，此篇對話的時數太長，處理起來有點不方便。

種類	breathe	clear throat	click	cough	hiccup	laugh	sign
個數	96	1	2	5	3	75	1
百分比	25.5	0.3	0.5	1.3	0.8	19.9	0.3
種類	silence	smack	sniffle	swallow	unrecognizable	pause	
個數	17	38	5	2	3	129	
百分比	4.5	10.1	1.3	0.5	0.8	34.2	

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 612)個數	75	230	98	27	4	4	7	6
百分比	12.3	37.6	16	4.4	0.7	0.7	1.1	1
particle 種類	嗎、嘛	吧	啦	厚	哈	囉	其他	
個數	72	16	25	9	1	3	35	
百分比	11.8	2.6	4.1	1.5	0.2	0.5	5.7	

因為對話時間長的緣故，所以此對話收錄了很多類型的雜音。

(6) 運動:

一男一女的對話。很自然的聊天，但女生有時會用抱怨的口氣，這一段男女生講話都有點斷斷續續結巴的感覺。當男生和女生討論起關於休息室的話題時，兩人講話都不太流暢，而且這一段女生常常會出現很大的呼吸聲。

種類	breathe	click	cough	laugh	silence	smack	swallow	
個數	95	2	1	9	2	2	7	
百分比	73.1	1.5	0.8	6.9	1.5	1.5	5.4	
種類	trill	unrecognizable	pause					
個數	1	6	5					
百分比	0.8	4.6	3.9					
particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢

(共 234)個數	17	88	3	59	1	19	15	3
百分比	7.3	37.6	1.3	25.2	0.4	8.1	6.4	1.3
particle 種類	嗎、嘛	吧	啦	厚	囉	其他		
個數	16	1	5	1	1	5		
百分比	6.8	0.4	2.1	0.4	0.4	2.1		

大部分的 breathe 都是來自錄音中的女生，其他雜音卻很少出現。

(7) 算命 05

一男一女的對話。是一段有內容、聽起來輕鬆的對話，內容談到算命以及生死、姻緣。男女講話口齒都很清晰，表達能力也不錯，互動的對談很多，有應有答，以致於標註的時候必須要注意聽兩人聲音重複的部份。

種類	breathe	click	cough	laugh	silence	smack	sniffle
個數	54	2	8	36	1	3	1
百分比	25.4	0.9	3.8	16.9	0.5	1.4	0.5
種類	swallow	unrecognizable	pause				
個數	3	8	97				
百分比	1.4	3.8	45.5				

particle 種類	喔	啊	ㄟ、世	嗯	哇	嗯哼	呃、兒	呢
(共 313)個數	32	91	19	50	2	61	1	9
百分比	10.2	29.1	6.1	16	0.6	19.5	0.3	2.9
particle 種類	耶	嗎、嘛	吧	啦	厚	哈	唉唷	其他
個數	2	14	8	8	8	3	3	2
百分比	0.6	4.5	2.6	2.6	2.6	1	1	0.6

此段錄音的 Unrecognizable speech sound 出現的次數頗多。

(8) 算命 06

一男一女的對話，男女講話口齒都很清晰，兩人的互動也很頻繁，是一段輕鬆的對話，表達能力也不錯，主要是在談論算命的一種另類方法，男生的特點是常常會清喉嚨。

種類	breathe	pause	cough	laugh	smack	sneeze	swallow	unrecognizable
個數	48	89	24	13	3	1	2	9
百分比	25.4	47.1	12.7	6.9	1.6	0.5	1	4.8

particle 種類	喔	啊	ㄟ、世	嗯	哇	嗯哼	呃、兒	呢
-------------	---	---	-----	---	---	----	-----	---

(共 309)個數	48	57	19	33	4	73	4	8
百分比	15.5	18.5	6.2	10.7	1.3	23.6	1.3	2.6
particle 種類	耶	嗎、嘛	吧	啦	厚	哈	唉唷	其他
個數	1	25	9	14	7	2	1	4
百分比	0.3	8.1	2.9	4.5	2.3	0.7	0.3	1.3

(9) 談狗狗

一男一女的對話。一開始男生和女生很開心的談論狗狗授精懷孕的事情，這一段男女重複說話較少，清晰好理解。再來女生談論她們家裡三隻小狗的出生情形時常有說話結巴的情形。最後來談論起三隻小狗的命名由來，這一段因為女生刻意模仿使得很多發音不正確聽不懂，不好標註。

種類	breathe	click	laugh	sign	silence	smack	swallow
個數	95	1	47	1	1	5	1
百分比	49.7	0.5	24.6	0.5	0.5	2.6	0.5
種類	unrecognizable	pause					
個數	1	39					
百分比	0.5	20.4					

particle 種類	喔	啊	ㄟ、世	ㄊ 嗯	嗯	哇	嗯哼	呃、兒
(共 230)個數	29	82	25	5	22	1	1	2
百分比	12.6	35.7	10.9	2.2	9.6	0.4	0.4	0.9
particle 種類	呢	嗎、嘛	吧	啦	厚	囉	其他	
個數	10	18	6	10	5	2	12	
百分比	4.4	7.8	2.6	4.4	2.2	0.9	5.2	

輕鬆的對談，所以笑聲與呼吸聲出現很多。

(10) 旅行與音樂 01

兩男的對話。開心的語調談論即將要去歐洲觀賞音樂表演。對話中常常出現義大利文以及法文的字彙和練習，打起來比較麻煩而且不容易辨認。男 1 的聲音很容易聽得清楚但不時的會唱起歌來，且常常發出一些怪聲和誇張的笑聲；男 2 的聲音較低沉，咬字比較不清楚，講話聲音忽大忽小，錄音過程中常做出像打嗝、敲打麥克風和打哈欠這些事情，也錄進很多吵雜的聲音。錄音檔中 silence 出現的機率也很多。

種類	breathe	pause	cough	hiccup	laugh	silence	unrecognizable	yawn
個數	45	201	4	2	20	28	12	2

百分比	14.3	64	1.3	0.6	6.4	8.9	3.8	0.6
-----	------	----	-----	-----	-----	-----	-----	-----

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 304)個數	60	86	33	28	1	2	1	5
百分比	19.7	28.3	10.9	9.2	0.3	0.7	0.3	1.6
particle 種類	嗎、嘛	吧	啦	厚	哈	唉唷	唉呀	其他
個數	38	10	22	1	6	3	4	4
百分比	12.5	3.3	7.2	0.3	2	1	1.3	1.3

Unrecognizable 的部分非常嚴重，主要是來自於男 1 發出的怪聲。

(11) 旅行與音樂 02

兩男的對話。很輕鬆在討論有關於音樂的雜誌。男 2 對音樂不是很了解，在對話中仍不斷發出怪聲以及種種的疑問，但是咬字很清楚；男 1 在聊到音樂的時候咬字就比較清楚，講的內容也讓人可以了解。因為過程中兩人有使用網路，以致於常常出現敲鍵盤的雜音。

種類	breathe	pause	hiccup	laugh	sign	silence	swallow	unrecognizable
個數	53	155	2	5	3	8	1	11
百分比	22.3	65.1	0.8	2.1	1.3	3.4	0.4	4.6

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	呃、兒	呢	嗎、嘛
(共 289)個數	102	58	14	36	5	6	5	29
百分比	35.3	20.1	4.8	12.5	1.7	2.1	1.7	10
particle 種類	吧	啦	厚	哈	其他			
個數	12	8	2	9	3			
百分比	4.2	2.8	0.7	3.1	1			

Unrecognizable 大部分來自敲鍵盤的雜音，次數還蠻多的。

(12) 旅行與音樂 03

兩男的對話。男 1 一開始在抱怨一些事情，這時候突然冒出一位女性和男 2 的對話，大約一分鐘結束。兩人有時候會發出些怪聲，中間男 1 不時的打嗝和打哈欠。當男 1 開始試聽網站上巴托利的演唱，標註過程中要很費力才能聽出來，因為音樂放得太大聲了。且兩人不時會念出英文、義大利文或法文，在標註的時候不容易分辨。且男 1 不時會唱起歌來，男 2 有時也會跟著唱和，而雜音也是不斷的出現，使這音檔的打字變得非常困難。

種類	breathe	hiccup	laugh	silence	smack	swallow	unrecognizable
----	---------	--------	-------	---------	-------	---------	----------------

個數	88	3	13	21	2	3	11
百分比	28.7	1	4.2	6.8	0.7	1	3.6
種類	yawn	pause					
個數	5	161					
百分比	1.6	52.4					

particle 種類	喔	啊	ㄟ、世	嗯	哇	嗯哼	呃、兒	呢
(共 262)個數	80	55	21	29	1	2	1	6
百分比	30.5	21	8	11.1	0.4	0.8	0.4	2.3
particle 種類	耶	嗎、嘛	吧	啦	厚	哈	其他	
個數	2	32	4	16	2	3	8	
百分比	0.8	12.2	1.5	6.1	0.8	1.2	3.1	

(13) 旅行與音樂 04

兩男的對話。背景音樂和雜音還是不斷出現。男 1 開心的介紹音樂家、音樂劇、和故事的背景，穿插著英文和中文。中間男 1 有時會讀出關於音樂的英文資料，兩人沒有什麼互動，但大致上來說兩人的口齒都蠻清晰的，但有時某人會大叫將另一個人的聲音蓋過去，使得標註過程變得比較緩慢，兩人也常發出怪聲和唱歌，是段沒有規劃內容的錄音。

種類	breathe	clear throat	cry	hiccup	laugh	silence	smack
個數	83	1	3	1	6	8	1
百分比	31.2	0.4	1.1	0.4	2.3	3	0.4
種類	swallow	unrecognizable	pause				
個數	4	26	133				
百分比	1.5	9.8	50				

particle 種類	喔	啊	ㄟ、世	嗯	哇	嗯哼	呃、兒	呢
(共 221)個數	68	55	12	16	4	1	3	12
百分比	32.2	26.1	5.7	7.6	1.9	0.5	1.4	5.7
particle 種類	嗎、嘛	吧	啦	厚	哈	唉呀	其他	
個數	21	4	7	3	1	1	3	
百分比	10	1.9	3.3	1.4	0.5	0.5	1.4	

Unrecognizable 非常的嚴重，主要是來自背景音樂和其他雜音。

(14) 旅遊 01

一男一女的對話。兩人很開心的談論出國的經驗，整段對話幾乎是由女生在

主導，男生發出 NHM 的回應的頻率很高，男女講話都還算清楚，也沒有什麼雜音出現，這一集兩人說話時出現的連音變多，不過大致上還是可以辨認出來講的是什麼字。

種類	breathe	clear throat	laugh	smack	unrecognizable	pause			
個數	79	6	23	1	5	29			
百分比	55.2	4.2	16.1	0.7	3.5	20.3			
particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢	
(共 351)個數	49	79	27	49	13	60	1	7	
百分比	14	22.5	7.7	14	3.7	17.1	0.3	2	
particle 種類	嗎、嘛	吧	啦	厚	哈	唉唷			
個數	26	1	20	13	5	1			
百分比	7.4	0.3	5.7	3.7	1.4	0.3			

(15) 旅遊 02

一男一女的對話。兩人很開心的談論出國逛街購物，女跟男分享她在國外的經驗，主要還是女在講話，男會常常發出 NHM 的聲音來回應，談話中會出現兩人同時大笑和講話的情況，比較不好打字，有時候女生講話有點快，而且越來越小聲，很難分辨出來她在講什麼。

種類	breathe	clear throat	laugh	smack	swallow	unrecognizable	pause
個數	92	4	34	2	1	9	56
百分比	46.5	2	17.2	1	0.5	4.6	28.3

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 377)個數	58	61	19	25	6	133	13	6
百分比	15.4	16.2	5	6.6	1.6	35.3	3.5	1.6
particle 種類	嗎、嘛	吧	啦	厚	哈	唉唷	唉呀	其他
個數	24	8	8	8	4	1	1	2
百分比	6.4	2.1	2.1	2.1	1.1	0.3	0.3	0.5

3.2.2 對話語料 3 中央社製播

編號	節目名稱/檔名	總長度	音節數	Speaker 數
1	上台報告經驗談	29分22秒	8581	2

2	五專生活	31分30秒	8685	2
3	養小狗	29分48秒	7328	2
4	狗狗命名	10分20秒	2812	2
5	香港人	8分22秒	2380	2
6	棒球	9分57秒	2795	2
7	週年慶	10分20秒	3087	2
8	雜談	10分20秒	2735	2
9	租屋經驗	10分31秒	3349	2
10	做菜	10分09秒	2800	2
11	電視劇	10分51秒	3214	2
12	吃點心談相親	10分03秒	2667	2
13	香辣蟹	19分11秒	4792	2
14	閒話家常	16分08秒	4158	2
15	節目討論	8分59秒	2637	2
16	漁人碼頭	13分33秒	3269	2
17	關東煮	9分29秒	2081	2

總數 17 個 總時間 248:53

節目性質與標註狀況說明：

(1) 上台報告經驗談

一男一女的對話。輕鬆自然的對話，兩人咬字都很不清楚，講話也都很含糊，隨便講講就帶過去了，有些地方也因為實在是太模糊以致於沒法辨認。講話時都會因為急於表達自己的想法，常常造成結巴和很大的呼吸聲以及砸嘴的聲音，中間不明的雜音也很多。

種類	breathe	cough	laugh	pause	smack	sniffle	swallow	unrecognizable
個數	401	1	54	286	32	1	1	41
百分比	49.1	0.1	6.6	35	3.9	0.1	0.1	5

particle 種類	喔	啊	ㄟ、世	嗯	哇	嗯哼	呃、兒	呢
(共 504)個數	66	130	82	90	2	12	5	13
百分比	13.1	25.8	16.3	17.9	0.4	2.4	1	2.6

particle 種類	嗎、嘛	吧	啦	厚	哈	唉唷	其他
個數	41	10	28	2	4	6	13
百分比	8.1	2	5.6	0.4	0.8	1.2	2.6

Unrecognizable 與 breath 還有 pause 都非常的嚴重。

(2) 五專生活

一男一女的對話。**很開心的對話**，兩人講話都很含糊，幾乎全部都是以連音的方式在講話，有的時候就算某些部份聽不清楚還是可以了解整句話的意思，可是有時候即使重複聽很多遍還是不能了解整句話在講什麼，尤其女講話容易結巴，句子中常常出現不必要的字，才能夠表達她所要講的東西。兩人時常大笑，或者是講話過於大聲，以致於不容易聽到另外一個人在講什麼。

種類	breathe	clear throat	cough	cry	hiccup	laugh	silence
個數	368	2	1	1	1	84	7
百分比	43.1	0.2	0.1	0.1	0.1	9.8	0.8
種類	smack	swallow	unrecognizable	pause			
個數	24	8	27	330			
百分比	2.8	0.9	3.2	38.7			

particle 種類	喔	啊	ㄟ、世	嗯	哇	嗯哼	呃、兒	呢
(共 584)個數	86	192	70	89	1	13	3	5
百分比	14.7	32.9	12	15.2	0.2	2.2	0.5	0.9
particle 種類	嗎、嘛	吧	啦	厚	哈	唉唷	其他	
個數	53	5	31	6	2	17	11	
百分比	9.1	0.9	5.3	1	0.3	2.9	1.9	

Unrecognizable 與 breath 還有 pause 非常的嚴重。

(3) 養小狗

一男一女的對話。**很開心的對話**，兩人講話口齒都很不清楚，在講一些常用字時會有省略的情況發生，雖然那些字不是重點但聽的時候就會造成困擾，兩人說話都會有點結巴，口語表達的能力不是很好，以致於意思表達出來了，但是用了非常多不完整的句子，發音也都因為結巴而不清楚，造成標註的困難性。女生在談話中常常**情緒會很激動**，發出奇怪的聲音，發音也都不是很正確，要聽上下文才能猜測出她說的是什麼字。男生常常是**一邊笑一邊講話**的，所以也不是很容易聽得出來在講什麼。

種類	breathe	clear throat	cough	cry	laugh	silence	smack
個數	338	9	2	1	31	16	31
百分比	44.5	1.2	0.3	0.1	4.1	2.1	4.1
種類	swallow	unrecognizable	pause				

個數	8	40	284
百分比	1.1	5.3	37.4

particle 種類	喔	啊	ㄟ、世	嗯	哇	嗯哼	呢	耶
(共 512)個數	56	168	58	64	3	11	2	1
百分比	10.9	32.8	11.3	12.5	0.6	2.1	0.4	0.2
particle 種類	嗎、嘛	吧	啦	厚	哈	唉唷	唉呀	其他
個數	52	15	35	4	3	21	2	17
百分比	10.2	2.9	6.8	0.8	0.6	4.1	0.4	3.3

Unrecognizable 非常的嚴重，主要是因為兩個人都口齒不是很清楚。

(4) 狗狗命名：

一對男女的對話。一開始男生批評的語氣說現在的流行歌歌詞很膚淺，但後來就是很開心的對談了。女生講話不太清楚，有時候會離開麥克風，或是含糊不清，不是很容易辨別。其中他們有提到別人的名字，但是都沒說的很清楚，很難分辨。

種類	breathe	cough	laugh	smack	pause
個數	10	1	14	6	86
百分比	8.5	0.9	12	5.1	73.5

particle 種類	喔	啊	ㄟ、世	嗯	嗯哼	呃、兒	呢	嗎、嘛
(共 218)個數	23	94	19	17	1	5	14	22
百分比	10.6	43.1	8.7	7.8	0.5	2.3	6.4	10.1
particle 種類	吧	啦	其他					
個數	7	7	9					
百分比	3.2	3.2	4.1					

Pause 出現比其他雜音還要高。

(5) 香港人

一男一女的對話。男生輕鬆的說到香港出差的經驗，還模仿起香港人說話的口音。這一段談話最主要是模仿廣東話的地方最難辨識，其它的都很清晰。

種類	breathe	cough	laugh	silence	smack	pause
個數	6	1	20	1	1	4
百分比	18.2	3	60.6	3	3	12.1

particle 種類	喔	啊	ㄟ、ㄝ	嗯	呃、兒	呢	耶	嗎、嘛
(共 247)個數	27	126	14	25	1	2	9	20
百分比	10.9	51	5.7	10.1	0.4	0.8	3.6	8.1
particle 種類	吧	啦	哈	其他				
個數	1	9	2	11				
百分比	0.4	3.6	0.8	4.5				

輕鬆的對談，所以笑聲出現較多。

(6) 棒球：

一男一女對話。主要是開心輕鬆的談論有關棒球事情。兩邊聲音都很清楚。很少有重複的聲音出現，聲音很清晰，且很好標註。

種類	breathe	laugh	smack	pause
個數	3	10	5	11
百分比	10.3	34.5	17.2	37.9

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 174)個數	25	88	11	7	3	1	2	9
百分比	14.4	50.6	6.3	4	1.7	0.6	1.1	5.2
particle 種類	嗎、嘛	吧	啦	其他				
個數	14	3	9	2				
百分比	8	1.7	5.2	1.1				

(7) 週年慶

一男一女的對話。很輕鬆的談論到百貨公司買東西的經驗。這一段談話男女生重複短詞的頻率頗高，例如：對對對對之類的，不太好處理。

種類	breathe	laugh	smack	pause
個數	14	15	6	1
百分比	38.9	41.7	16.7	2.8

particle 種類	喔	啊	ㄟ、ㄝ	嗯	嗯哼	呃、兒	呢	耶
(共 246)個數	46	93	27	14	4	2	8	1
百分比	18.7	37.8	11	5.7	1.6	0.8	3.3	0.4
particle 種類	嗎、嘛	吧	啦	厚	囉	其他		
個數	16	14	9	1	4	7		

百分比	6.5	5.7	3.7	0.4	1.6	2.8
-----	-----	-----	-----	-----	-----	-----

(8) 雜談

一男一女的對話。男生的聲音都算平穩，女生的聲音則比較有起伏，屬於比較開心的語調。後來在談到小朋友似乎不懂禮貌的地方，男生的聲音就比較憤慨。後段的地方男生用廣東話的語調在講話。

種類	breathe	laugh	sign	smack
個數	1	12	1	4
百分比	5.6	66.7	5.6	22.2

particle 種類	喔	啊	ㄟ、世	嗯	呃、兒	呢	嗎、嘛	吧
(共 236)個數	44	113	17	12	2	7	14	7
百分比	18.6	47.9	7.2	5.1	0.8	3	5.9	3

particle 種類	啦	囉	其他
個數	9	2	9
百分比	3.8	0.8	3.8

(9) 租屋經驗：

一男一女的對話。很輕鬆的談論租屋的經驗，男生的聲音很清楚，不過因為女生常常笑，而且發出阿等驚奇的叫聲，讓人聽不清楚，而且其中有一段男跟女兩人都搶著要講話，所以會很混亂，無法辨認。

種類	breathe	clear throat	laugh	silence	smack	pause
個數	14	1	43	1	1	67
百分比	11	0.8	33.9	0.8	0.8	52.8

particle 種類	喔	啊	ㄟ、世	嗯	嗯哼	呃、兒	呢	嗎、嘛
(共 247)個數	51	56	22	43	1	14	16	24
百分比	20.6	22.7	8.9	17.4	0.4	5.7	6.5	9.7

particle 種類	吧	啦	厚	其他
個數	4	8	1	7
百分比	1.6	3.2	0.4	2.8

笑聲出現很多，大多來自於女生的部份。

(10) 做菜：

一男一女的對話。主要是開心的談論做菜的經驗，整體上講話很清楚，不過

有時候要急於表達自己的意見，會說的比較快比較激動，這時候通常聽不太清楚，像”可是”我們平常講話都會很快的帶過去，也不會去刻意去強調這兩個字，畢竟少了這兩個字還是可以知道對方的意思，但是變造成了標註上的困難。

種類	breathe	clear throat	cough	laugh	pause
個數	17	2	1	30	36
百分比	19.8	2.3	1.2	34.9	41.9

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 246)個數	30	101	4	32	2	7	12	11
百分比	12.2	41.1	1.6	13	0.8	2.8	4.9	4.5
particle 種類	嗎、嘛	吧	啦	厚	囉	其他		
個數	20	7	15	1	1	3		
百分比	8.1	2.8	6.1	0.4	0.4	1.2		

雙方有時候會急於表達自己意見，導致 pause 出現頗高。

(11) 電視劇：

一男一女的對話。很開心的討論華語、台語及日本的電視劇。男生講話是很標準，不過女生太常笑，以致於話常常沒講清楚。還有裡面常會提到一些人名，又說的很快，很難判別。裡面還說很多的台語，而且他們自己也很不確定是不是講對，所以更加模糊難以判別。

種類	breathe	clear throat	laugh	smack	pause
個數	16	1	24	7	79
百分比	12.6	0.8	18.9	5.5	62.2

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	呃、兒	呢	嗎、嘛
(共 258)個數	38	61	14	68	2	23	7	14
百分比	14.7	23.6	5.4	26.4	0.8	8.9	2.7	5.4
particle 種類	吧	啦	厚	囉	其他			
個數	4	10	14	1	2			
百分比	1.6	3.9	5.4	0.4	0.8			

(12) 吃點心談相親

兩女的對話。一開始是很開心的討論改名西，但說的零零碎碎的，從 53 秒到 3 分 50 秒的。之後兩人用批評的口氣在女星造型身上的。

種類	breathe	pause	cough	hiccup	laugh	sign	silence	smack
個數	12	2	3	1	32	4	6	5
百分比	18.5	3.1	4.6	1.5	49.2	6.2	9.2	7.7

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呢	耶
(共 186)個數	21	74	9	46	1	2	3	2
百分比	11.3	39.8	4.8	24.7	0.5	1.1	1.6	1.1

particle 種類	嗎、嘛	吧	啦	厚	其他
個數	8	6	6	3	5
百分比	4.3	3.2	3.2	1.6	2.7

很開心的對談，所以笑聲比其他雜音多。

(13) 香辣蟹：

是兩女的對話。他們開心的邊吃東西邊講話，所以很難聽清楚，而且又有很多雜音，兩個女生聲音很像，很難辨視標註。

種類	breathe	clear throat	laugh	silence	smack	pause
個數	22	3	61	5	5	128
百分比	9.8	1.3	27.2	2.2	2.2	57.1

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 319)個數	33	123	30	63	4	3	6	4
百分比	10.3	38.6	9.4	19.7	1.3	0.9	1.9	1.3

particle 種類	嗎、嘛	吧	啦	厚	其他
個數	29	11	6	1	6
百分比	9.1	3.4	1.9	0.3	1.9

因為邊吃邊說話的緣故，所以 pause 出現頗高。

(14) 閒話家常

兩女的對話。很愉悅的對話。女 1 很喜歡嗯嗯的附和所以造成常會需要特別打重疊談話。女 2 說話喜歡含在嘴裡講，讓人常常聽不清楚語意。兩人都喜歡插入台語說話，這個部分需要特別區分出來，造成標註的困難。

種類	breathe	pause	cough	hiccup	laugh	sign	silence	smack
個數	21	2	4	2	37	4	3	5
百分比	26.9	2.6	5.1	2.6	47.4	5.1	3.8	6.4

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 343)個數	33	102	19	124	4	1	2	15
百分比	9.6	29.7	5.5	36.2	1.2	0.3	0.6	4.4
particle 種類	嗎、嘛	吧	啦	厚	其他			
個數	18	6	6	3	10			
百分比	5.2	1.7	1.7	0.9	2.9			

(15) 節目討論

兩女的對話，因為是節目開始前的討論，**小聲的語氣**，所以很多話講的都很**不清楚**，也不斷的重複講一些已經講過的話，而且這兩個女生的聲音的特色不是很明顯，所以非常難分辨，很多地方是沒辦法確定是誰說的，尤其當兩人同時講話的時候，聲音幾乎是重疊在一起的，無法聽出哪一個人講了什麼，只能靠上下文來猜測，整段對話幾乎者沒有辦法百分之百確定內容，只有在一開始模擬廣播時女 1 的聲音非常清楚，其餘時間兩人講話的聲調都很相近、講話內容也都模糊不清。

種類	breathe	pause	cry	laugh	silence	smack	unrecognizable	yawn
個數	68	98	1	14	1	5	16	2
百分比	33.2	47.8	0.5	6.8	0.5	2.4	7.8	1

particle 種類	喔	啊	ㄟ、ㄝ	嗯	嗯哼	呃、兒	呢	嗎、嘛
(共 122)個數	4	34	8	27	7	7	12	11
百分比	3.3	27.9	6.6	22.1	5.7	5.7	9.8	9
particle 種類	吧	啦	厚	唉唷	其他			
個數	4	4	2	1	1			
百分比	3.3	3.3	1.6	0.8	0.8			

Unrecognizable 的部份很多，主要是因為節目開始前語調很小的關係。

(16) 漁人碼頭

兩女的對話。兩人講話口齒都算蠻**清楚且很開心**的，對話中常常出現兩人的笑聲，有時候會蓋住另一個人說話的聲音以致於無法辨認。因為是兩個女聲，所以花在分辨聲音的時間上比較多。特殊情況是女 2 習慣講「結果」這個詞，但音都連在一起了無法確定她在講什麼，一開始需要時間重複聽，後來也就慢慢習慣了，算是蠻輕鬆的打字過程。

種類	breathe	pause	clear throat	cough	laugh	smack	swallow	unrecognizable
個數	105	123	1	2	100	6	3	9
百分比	30.1	35.2	0.3	0.6	28.7	1.7	0.9	2.6

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 367)個數	43	93	12	124	3	26	5	15
百分比	11.7	25.3	3.3	33.8	0.8	7.1	1.4	4.1

particle 種類	嗎、嘛	吧	啦	厚	哈	其他
個數	11	5	20	2	5	3
百分比	3	1.4	5.4	0.5	1.4	0.8

Breathe 與 pause 和 laugh 非常嚴重。

(17) 關東煮

是兩個女生的談話。兩個女生都屬於開心的語調，但是女二在談論食物的部分口氣抱怨。但是後來兩個女生吃起東西，講話會有點含糊。而且背景會一直有食物包裝的雜音。

種類	laugh	smack	swallow
個數	28	3	1
百分比	87.5	9.4	3.1

particle 種類	喔	啊	ㄟ、ㄝ	嗯	哇	嗯哼	呃、兒	呢
(共 177)個數	30	54	10	30	2	1	1	10
百分比	16.9	30.5	5.6	16.9	1.1	0.6	0.6	5.6
particle 種類	耶	嗎、嘛	吧	啦	厚	其他		
個數	1	13	8	9	1	7		
百分比	0.6	7.3	4.5	5.1	0.6	4		

整段對話是開心的語調，所以笑聲出現比較多。