

行政院國家科學委員會補助專題研究計畫成果報告

利用語音辨識技術建立可搜尋之耳鼻喉科教學數位影音資料庫

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 90-2511-S-038-001-

執行期間： 2001 年 8 月 1 日至 2002 年 7 月 31
日

計畫主持人：李飛鵬

共同主持人：李友專 徐建業 林凱南

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：臺北醫學大學/臺北醫學大學附設醫院耳鼻喉科

中 華 民 國 91 年 10 月 1 日

行政院國家科學委員會專題研究計畫成果報告國科會

利用語音辨識技術建立可搜尋之耳鼻喉科教學數位影音資料庫

Building a Searchable Digital Video Database for Otolaryngology Teaching by using
Speech Recognition Technology (2-2)

計畫編號：NSC 90-2511-S-038-001-

執行期限：2001年08月01日至2002年07月30日

主持人：李飛鵬 臺北醫學大學/臺北醫學大學附設醫院

共同主持人：李友專 臺北醫學大學醫學資訊研究所

共同主持人：徐建業 臺北醫學大學醫學資訊研究所

共同主持人：林凱南 臺灣大學耳鼻喉科部

計畫參與人員：盧建宏 臺北醫學院醫學資訊研究所

計畫參與人員：林明錦 臺北醫學院醫學資訊研究所

中文摘要

本計畫主要之目的在於利用語音辨識技術，將耳鼻喉科在教學影片中的語音轉換成文字。配合本計劃所開發之軟體，將文字標註時間碼，利用網路及多媒體資料庫技術，我們可以藉由搜尋特定文字來觀賞影片段落，如此一來學生及醫師即可利用網路快速，不受地點及時間限制，來學習耳鼻喉科教學。

Abstract

The purpose of this project is to create and integrate multimedia medical resource for medical community. In this project, we realize the "speech to text" technology by voice recognition. In this project, the user can access any section of video clip randomly by searching the time code based text. Due to this, the user who using this system can watch the movie clips more efficiently.

Keywords

Otorhinolaryngology, voice recognition, Digital Video, WWW

二、緣由與目的

隨著數位科技的日益成熟，越來越多的影音資料，在我們的生活中扮演了越來越重要的角色。也因此，人們也嘗試著利用這許多資料，來作各方面的應用。首先遇到的問題，就是如何做索引及搜尋的工作。日益龐

雜的影音資料，甚至是 3D 立體模型，比起過去的文字檢索，其複雜度更是不可同日而語。而手術錄影帶是專科醫學教育一項重要的資產，傳統的錄影帶都是以類比方式錄影，播放則是線性方式從頭開始到尾。近年來由於數位媒體儲存的進步，影音資料有從類比轉成數位，從線性轉成非線性的趨勢。配合寬頻網路及影音伺服器的架設，可使多人多處同時觀看錄影內容，達到專科醫學遠距再教育的目的。

所謂非線性的播放方式即是可搜尋之影音資料，使用者可依喜好找尋需要之錄影片段，不需從頭到尾把錄影帶看一遍。這種技術在非醫學方面有個很好的例子，就是在柯林頓誹聞案大審判時的錄影內容，Virage公司和 AltaVista 合作，成功的運用 VideoLogger 軟體製作成可搜尋之錄影內容，使用者可以文字尋找某一特定的錄影片段，這是第一個在網路可搜尋之影片。

耳鼻喉科手術由於視野狹小，故常常使用顯微鏡或內視鏡，接至錄影設備後便可將手術過程完整的紀錄下來。常見的手術如耳部的鼓室成型術 (tympanoplasty)，鼻部的功能性鼻竇內視鏡手術 (functional endoscopic sinus surgery)，喉部的喉頭直達鏡顯微手術 (laryngomicroscopic surgery)。以鼓室成型術 (tympanoplasty) 為例，一台手術通常需要

兩個小時，如果使用者只對聽小骨中的磴骨 (stapes)處理有興趣，如何不需從頭找起呢？

手術錄影帶是醫學教育一項重要的資產，傳統的錄影帶都是以類比方式錄影，播放則是線性方式從頭開始到尾，這些影片在尋找使用上非常不便，而且影帶會隨著時間耗損，無法永久保存。近年來由於數位媒體儲存的進步，影音資料有從類比轉成數位，從線性轉成非線性的趨勢，而這些非線性的影音資料的 index 及 search 便成為一個熱門的課題。

結果與討論

本研究計畫之目的在研究及發展一可檢搜索尋及互動之耳鼻喉科手術數位影音資料庫；並結合網際網路，建立一影音伺服器，具備多媒體互動教學功能，以協助耳鼻喉科醫師再進修或住院醫師訓練，提昇醫療的品質。在第二階段的計劃中，我們依循 MPEG-7 標準，建立符合國際交換標準的多媒體資料庫，另外則是建構一 Story Telling Platform (STP)，來負責讀取影片中 MPEG-7 (Scenarios) 腳本資料及第一年計劃的關鍵字資料庫，透過網際網路的瀏覽器的介面，提供互動式的教學影片。這計劃完成兩個部分：

I. **MPEG-7 Encoder/Decoder engine:** MPEG-7 也可稱為 (Multimedia Content Description Interface)，其中對多媒體資料有相當廣泛的定義，MPEG-7 採用 XML Schema Language as the MPEG-7 Description Definition Language (DDL) 本計劃將對原本之耳鼻喉科影音資料庫，主要做兩方面的描述與紀錄。我們選使用了 Spoken Content description tools，我們將第一階段計劃之資料，將演講者的演講內容儲存起來，提供索引及藉此連結相關的多媒體物件。目前我們採用的資料庫格式為 Microsoft SQL 2000，此資料庫支援 XML 的輸出及輸入，第一階段計劃中我們已完成 Information Retrieval 的引擎，在第二階段

計劃中，以 MPEG-7 為 Coding Scheme，利用 Microsoft SQL 2000，及 Active Server Page 技術完成 MPEG-7 Encoder/Decoder engine。

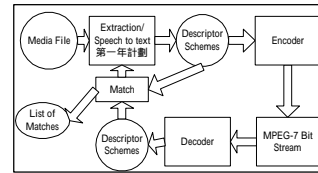


Figure1. The implementation of MPEG-7 Encoder/Decoder engine

I. **II. Story-Telling Platform (STP):**在影片 MPEG-7 Metadata 中我們可以儲存(keyword, caption, 及 Event)，我們提出以網際網路瀏覽器為基礎平台的架構 Story-Telling Platform(STP, Figure.2)，其架構分述如下：

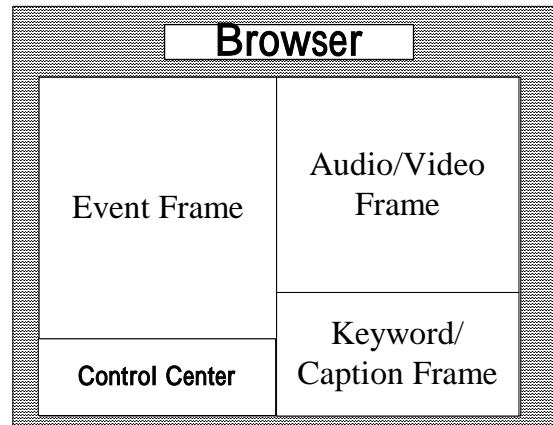


Figure.2 The framework of Story-Telling Platform (STP)

- Browser:以網際網路為基礎的瀏覽器
- Event Frame:以 HTML 為基礎,依據 Control Ceter 的控制顯示資料,由於 HTML 有發展良好的多媒體技術,所以此處的 Event 可以為 2D image, 3D model, audio-video data, 並保有 HTML 的互動性。
- Audio/Video Frame:目前採用 Streaming technology 播放影片或聲音。
- Keyword/caption frame:負責接收 Control Center 傳來的文字,此處的文字亦可提供作為 Searchable video 的關鍵字,可以做超連結。
- Control Center: 負責讀取影片的 Mpeg-7 format Metadata, 取得影片中的 Keyword/caption/Event, 並且在預設的 Timestamp 送出事件至 Event/Keyword/Caption frame, 並且也要控制影

片的播放。

而目前 MPEG-7 是一個相當新的標準，所以並沒有許多的軟體支援，所以我們必須開發如 STP 的環境來讀取額外影片的資訊，也增加系統開發的難度。不過，由於其強大具彈性的語法結構，我們還是願意採行它的標準，至少在跨網際網路的搜尋及館際的交換資料時，有一共通的平台。

計畫成果自評

本計畫之目的在於建立可搜尋耳鼻喉科的多媒體教學網站，我們已成功的將語音辨識軟體，應用在耳鼻喉科的手術錄影帶之中，將醫學的專用名詞辨識出來，並配合時間碼的標定，可以藉由 speech to text 的功能，搜尋文字來定位影片。

在今年度計劃中，我們共完成了：

1. 剪接製作咽喉部內視鏡影片共十六段約兩小時的影片。
2. 配合內視鏡影片，製作教學網頁

此計畫相關結果，也發表於 2002 年 AMIA 2002 Annual Symposium 本計劃團隊完成了第二年所預期的工作目標，並有論文之發表，完全符合計劃書中的預估完成工作成果。本計劃所開發之軟體與論文也將放置於網站 (<http://ENT.TMU.EDU.TW>)，公開於學術研究使用。

Reference

1. Brazma A, Hingamp P, Quackenbush J, Sherlock Gavin, et al. Minimum information about a micro array experiment – toward standards for microarray data. Nature Genetics (December, 2001):pp.365-371.
2. Digital Imaging and Communications in Medicine (DICOM) Version 3.0. American College of Radiologists / National Electrical Manufacturers Association (1993).
3. Hsu CY, Chen PH, Lee PS, Lin MC, Li YC.

Development of a Digital Museum for Human Body. Proceeding of the 3rd International Conference of Asian Digital Library (December, 2000): pp.267-274.

4. Lagoze C, Van de Sompe H. The Open Archives Initiative: Building a low-barrier interoperability framework. Proceeding of the ACM/IEEE Joint Conference on Digital Libraries (June, 2001): pp.54-62.
5. Li YC. Toward a medical information collective: trends in the development of digital libraries in medicine. Yearbook of Medical Informatics 2001(2001): pp.77-82