

## 本期要目

- |                           |        |
|---------------------------|--------|
| 壹. 2006 年國際會議預告           | 第二~五頁  |
| 貳. 團體會員入會優惠說明             | 第六~七頁  |
| 參. 專文- 語音辨識在矯正發音上的應用(周福強) | 第八~十二頁 |

### 賀！許聞廉教授當選 IEEE Fellow

中研院資訊所研究員許聞廉教授為本會現任理事，IEEE 肯定許聞廉教授在 Natural Language Systems and Bioinformatics 的傑出研究成果及貢獻，授予許聞廉教授 IEEE Fellow 之殊榮。特此恭賀！

### 賀！鄭秋豫教授當選

### Oriental-COCOSDA 召集人

COCOSDA (The International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques) 成立宗旨是促進國際間在口語處理研究基礎建設的互動與合作，而 Oriental COCOSDA 為其亞洲的區域組織，參與國家及地區包括台灣、大陸、香港、日本、韓國、新加坡、泰國、印度、印尼等。Oriental-COCOSDA 是 COCOSDA 最活絡的區域組織之一，中研院語言所研究員鄭秋豫教授獲得各國與會代表的一致支持，出任召集人，非常難得。特此恭賀！

### 「國語連續數字」資料庫開放申請

「國語連續數字」資料庫是由中華電信研究所錄製，智慧財產權為中華電信研究所所有，授權本會發行與國內學術研究使用。

本資料庫屬於麥克風朗讀語音，內容為不同長度之國語連續數字，由 100 人錄製而成，其中包括男、女性語者各 50 位。

語料庫內容：

1. 本語料庫內容為不同長度之國語連續數字。

2. 語者總數為 100，其中男、女性語者各 50 位。
3. 每位語者錄製 6 遍，每遍之語料內容均不相同。

檔案資料：

1. 本語料庫共分六個 SET，每個 SET 有 100(0~99) 位語者，每位語者有 40(0~39) 句，每句存為一個檔案。
2. 檔案名稱之格式為 dd\_tt.vat，其中 dd 表示語者編號，tt 表示句數編號。
3. 檔案之儲存格式主要係參考 MAT 語料之儲存格式，僅拼音碼部份是以注音符號表示；各檔案均包含 256 Bytes 的檔頭(header)，相關的取樣頻率、語者性別、處理時間、BIG5 碼與對應的注音符號均存於 header 中，詳細格式請參見 MAT file format。

擬申請者請參閱本會網頁：

([http://www.aclclp.org.tw/use\\_mat\\_c.php#cht1](http://www.aclclp.org.tw/use_mat_c.php#cht1))

### 漢語對話語音語料庫開放申請

「中央研究院現代漢語對話語音語料庫 (The Academia Sinica Mandarin Conversational Dialogue Corpus)」, 簡稱「漢語對話語音語料庫」(Sinica MCDC) 包含八個現代漢語對話的語音檔 (約 4.7GB) 與其語言內容之文字轉記檔 (約十三萬字)。所記錄的是 16 個語者在台灣使用的現代漢語口語 (錄音時間 2001 年三月至七月)。文字轉記檔由 TransList 介面輔助產生，包含每個說話輪在所屬語音檔裡的時間碼。本資料庫原始內容之智財權由中央研究院持有。公開授權資料分別以 wav 檔案以及純文字格式儲存。此中央研究院現代漢語對話語音語料庫得到中央研究院授權由本會發行。

擬申請者請參閱本會網頁：

([http://www.aclclp.org.tw/use\\_mat\\_c.php#mcdcl](http://www.aclclp.org.tw/use_mat_c.php#mcdcl))

## *2006 International Conferences List*

### **GAW-2006**

3rd International Wordnet Conference

Conference date: January 22-26, 2006

Location: Jeju Island, Korea

[http://www.globalwordnet.org/gwa/gwa\\_conf\\_03.htm](http://www.globalwordnet.org/gwa/gwa_conf_03.htm)

### **ICASSP-2006**

The 31st International Conference on Acoustics, Speech, and Signal Processing

Conference date : May 14-19, 2006

Location: Toulouse, France

<http://www.icassp2006.org/default.asp>

### **HLT/NAACL-2006**

Human Language Technology conference/North American chapter of the Association for Computational Linguistics annual meeting 2006

Conference date : June 4-9, 2006

Location: New York, USA

<http://nlp.cs.nyu.edu/hlt-naacl06/>

### **COLING/ACL-2006**

21<sup>st</sup> International Conferences on Computational Linguistics

Conference date: July 17-21, 2006

Location: Sydney, Australia

Deadline for submissions: February 28, 2006

<http://www.acl2006.mq.edu.au/>

### **ACM SIGIR-2006**

The 29th Annual International ACM SIGIR Conference

Conference date : August 6-11, 2006

Location: Seattle, Washington, USA

Deadline for submissions: January 30, 2006

<http://www.sigir2006.org/>

### **ICSLP-2006**

The Ninth International Conference on Spoken Language Processing

Conference date: September 17-21, 2006

Location: Pittsburgh PA, USA

Deadline for submissions: April 07, 2006

<http://www.interspeech2006.org/index.html>

# COLING/ACL 2006 Call for Papers

**21st International Conferences on Computational Linguistics**  
**July 17-21, 2006 Sydney, Australia**  
**<http://www.acl2006.mq.edu.au/>**

Following on from the highly successful events at Stanford (1984) and Montreal (1998), the International Committee on Computational Linguistics and the Association for Computational Linguistics will host their third joint conference in Sydney, July 17th-21st, 2006.

Papers are invited on substantial, original, and unpublished research on all aspects of computational linguistics, including, but not limited to: pragmatics, semantics, syntax, grammars and the lexicon; phonetics, phonology and morphology; lexical semantics and ontologies; word segmentation, tagging and chunking; parsing, generation and summarization; language modeling, spoken language recognition and understanding; linguistic, psychological and mathematical models of language; information retrieval, question answering, information extraction and text mining; machine learning for natural language; corpus-based modeling of language, discourse and dialogue; multi-lingual processing, machine translation and translation aids; multi-modal and natural language interfaces and dialogue systems; applications, tools and language resources such as annotated corpora; and system evaluation.

We are particularly interested in bridging the gap to other related communities, and so we especially welcome papers on speech technology and information retrieval. Also, for this first joint COLING/ACL conference to be held in Asia we especially encourage papers that address issues specific to the processing of Asian languages.

## **Categories of Papers**

Papers can be submitted to one of two categories: the REGULAR PAPER category or the POSTER category. Authors must designate one of these categories at submission time. Although both categories of submission will be the same maximum length, regular papers will be presented a single time to a potentially large audience during one of the parallel paper sessions at the conference, and poster presentations will be repeated several times before small groups of people at one of the conference poster sessions. Regular papers are most appropriate for presenting substantial research results, while posters are more appropriate for presenting an ongoing research effort. Regular papers and poster papers will appear in separate volumes of the proceedings.

## **Deadlines**

Paper submission deadline:	February 28, 2006
Notification of acceptance:	April 28, 2006
Camera ready papers due:	May 26, 2006
COLING/ACL 2006 Conference:	July 17-21, 2006

# ACM-SIGIR Call for Papers

**29th Annual International ACM SIGIR Conference in Research and Development in  
Information Retrieval**

**August 6-11, 2006 Seattle, Washington, USA**

**<http://www.sigir2006.org/>**

## Areas of Interest

SIGIR 2006 welcomes contributions related to any aspect of IR, but the major areas of interest are listed below. For each general area, two or more area coordinators will guide the reviewing process.

- Information Retrieval Theory and Models: User/Task-based IR theory
- IR Performance and Platforms: Performance, Compression, Scalability, Architectures, Mobile applications.
- Document and Query Representation: Text content representation (Indexing), Structure-based representation, XML, Metadata, Request representation, Queries, Summarization, Natural language processing for representation.
- IR Evaluation: Test collections, Evaluation methods and metrics, Experimental design, Data collection and analysis methods.
- User-oriented IR: Interactive IR, User interfaces and visualization, User studies, User models, Task-based IR.
- Web IR and Digital Libraries: Web IR, Intranet/enterprise search, Citation and link analysis, Adversarial IR, Digital libraries, Distributed IR.
- Cross-language Retrieval: Cross-language retrieval, Multilingual retrieval, Retrieval in languages other than English, Machine translation for IR.
- Multi-media Retrieval: Video and image retrieval, Audio and speech retrieval, Music retrieval.
- Document Tracking and Filtering: Topic detection and tracking, Content-based filtering, Collaborative filtering, Agents, Routing, Email spam.
- Question Answering and Extraction: Question answering, Information extraction, Lexical acquisition.
- Text Data Mining and Machine Learning for IR.
- Text Categorization and Clustering.
- Domain Specific IR Applications: Genomic IR, IR in software engineering, IR for chemical structures.

## Important Dates

November 14, 2005.	Mentoring Program deadline
January 30, 2006.	Full Research Papers due
February 27, 2006.	Poster, Demonstrations, Tutorial, Workshop, and Doctoral Consortium submissions due
April 10, 2006.	Notification of acceptance for all submissions

On the due dates, submissions will remain open until 17:00 PST.

# ICSLP-2006 Call for Papers

**The Ninth International Conference on Spoken Language Processing**  
**September 17-21, 2006 Pittsburgh PA, USA**  
**<http://www.interspeech2006.org/index.html>**

Interspeech 2006 will cover all aspects of speech science and technology. The conference will include plenary talks by world-class experts, tutorials, exhibits, special sessions covering interdisciplinary topics and important new emerging areas of interest, and parallel oral and poster sessions.

## **Topics of interest for paper submissions include:**

Linguistics, Phonology and Phonetics	Language/Dialect Identification
Prosody	Speech Recognition & Understanding
Paralinguistic & Nonlinguistic Information in Speech	Speaker Characterization & Recognition
Discourse & Dialogue	Multi-modal Processing
Speech Production	Multi-lingual Processing
Speech Perception	Spoken Language Information Retrieval
Physiology & Pathology	Spoken Language Translation
Language Acquisition, Development, & Learning	Spoken Dialogue Systems
Signal Analysis & Processing	Resources & Annotation
Acoustic Signal Segmentation & Classification	Assessment & Standards
Single- & Multi-channel Speech Enhancement	Education
Speech Coding & Transmission	Technologies for Language Learning
Spoken Language Generation & Synthesis	Technologies for the (Aged and) Challenged Applications
	Interdisciplinary Topics in Speech and Language

The Call For Papers flyer will appear here shortly.

## **Important Dates**

Proposals for Tutorials due	13th January 2006
Notification for Tutorials	1st February 2006
Four-page full papers due	7th April 2006
Notification for main papers	9th June 2006
End of early registration	23rd June 2006
Tutorial Day	17th September 2006
Main Conference dates	18-21st September 2006

# 中華民國計算語言學學會

## 團體會員入會優惠說明

### 一. 學會簡介：

(一). 會員結構：本會至目前為止共計有 192 名之會員，依身份別可分為團體會員 10 個、終身會員 99 位、個人會員 27 位及學生會員 56 位；其中國內會員佔 174 位及國外會員則有 18 位，茲依身份別之分類說明如下：

1. 團體會員：共計 10 個會員，分別為：致遠科技（股）公司、遠見科技（股）公司、工研院電通所、中研院資訊所、中研院詞庫小組、中研院計算中心、台灣微軟（股）公司、台達電子（股）公司、無敵科技（股）公司及香港中文大學電子工程所。其中香港中文大學兼具「團體」及「終身」會員之身分。
2. 終身會員：共計 99 位，其中國內佔 84 位，國外佔 15 位。會員之身分包含了國內外大學之教授（約佔 80%）、研究機構之研究人員（約佔 10%）及企業界高級主管（約佔 10%）。國外會員則遍及美國、日本、中國、香港及新加坡等地。
3. 個人會員：共計 27 位，其中國內佔 25 位，國外則佔 2 位。會員之身分包含了國內外大學之教授、研究機構之研究人員及企業界高層主管。
4. 學生會員：共計 56 位，加入本會之學生多為台灣大學、師範大學、交通大學、清華大學及成功大學等之碩博士生，就讀之科系包含資訊工程系、語言學研究所、電機工程系、資訊應用系、圖書資訊系及教育心理系。

### (二). 學術活動：

1. 年度大會：本會固定於 8~9 月份舉辦年度大會「自然語言暨語音訊號處理研討會」(ROCLING)，會議為期二天，與會者多為國內計算語言學界頂尖之學者、優秀之學生及相關領域之企業研究人員，參與人數每年約 200 人。
2. 國際會議：本會致力於計算語言學領域學術研究之發展及國際交流，因此，每年皆極力爭取主辦國際會議。近三年內本會曾主辦過之會議有：2002 年「第十九屆國際計算語言學研討會」(COLING-2002)、2002 年「第三屆國際中文口述語言處理研討會」(ISCSLP-2002)及 2005 年「第十九屆亞太地區語言、訊息暨計算國際研討會」(PACLIC-19)等。參與者皆超過 200 人次以上，尤以 COLING-2002 參與人數更高達 500 人次，國外學者佔了三分之二。
3. 專題討論會：本會每年定期舉行有關「語音訊號處理」、「資訊檢索」及「電腦輔助教學」研討會，以期達到研究成果交流及資源共享之目的。每次參加人數達 120 人。

### (三). 語料庫及出版品：

1. 本會除了定期發行《中文計算語言學期刊》及《通訊》外，並持續與計算語言學領域之學術研究單位，例如：中研院資訊所、中研院語言所、台灣大學、交通大學、清華大學及成功大學等單位進行合作，並取得授權發行「資料庫」及「技術報告」。
2. 本會授權發行重要之語料庫：平衡語料庫、中文句結構樹、國語語音資料庫 (Mat-2000)。
3. 近年本會授權發行之語料庫有：中文資訊檢索標竿測試集、中英雙語詞網、台灣口音英語語料庫、漢語對話語音資料庫及國語連續數字資料庫。

### (四). 參與國際會議組織：

1. 亞洲自然語言處理聯盟(AFNLP)：該聯盟之成員為亞洲各國之學會所組成，本會亦為成員之一，本會之代表人陳信希教授持續代表學會參與該組織所舉辦活動，藉以不斷吸收新資訊及進行國內外研究成果之交流。
2. 國際語音處理學會(ISCA)：為促進與該學會學員間之互動及研究成果之交流，本會

終身會員李琳山教授（台灣大學電機系），也是 ISCA 委員之一，促成本會與 ISCA 結為姊妹會，並推動雙會員制之優惠方案，凡同時加入 ISCA 及本會之個人會員者可享 25%的會費折扣。

3. **Oriental-COCOSDA** : The International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques 成立宗旨是促進國際間在口語處理研究基礎建設的互動與合作，而 **Oriental COCOSDA** 為其亞洲的區域組織，參與國家及地區包括台灣、大陸、香港、日本、韓國、新加坡、泰國、印度、印尼等。**Oriental-COCOSDA** 是 **COCOSDA** 最活絡的區域組織之一，本會前理事鄭秋豫教授出任為該組織之召集人，將有助學會與 **Oriental- COCOSDA** 學術之交流，促進學會的國際化。

## 二.入會獨享權益：

(一). 刊物及資料庫之優惠：

1. 贈閱中文計算語言學期刊(一年發行四期)：二本。
2. 贈閱通訊（一年六期）：二份。
3. 購買資料庫：可享會員優惠價。
4. 購買技術報告：可享會員優惠價。

(二). 由本會主辦之學術活動之優惠可享註冊費優惠五名名額。

(三). 提供服務：

1. 免費代為發送徵才啟事：透過 E-mail 方式發送與本會會員
2. 免費刊登廣告：免費於本會通訊刊登廣告一年二次，黑白版面，頁數二頁。

## 三.加入辦法：

(一). 填寫入會申請書。

(二). 年費：每年貳萬元整（終身團體會員二十萬元）

(三). 繳費方式：

匯款：劃撥戶名「中華民國計算語言學學會」，帳號：19166251

支票：支票抬頭請加註「中華民國計算語言學學會」。

# 語音辨識在矯正發音上的應用

周福強

銘傳大學資訊傳播工程系

## 1. 前言

在進入文章的主題之前，我想先談一些個人的經驗，因為這樣也許能讓大家更瞭解接下來要說明的內容，以及為什麼要談這個題目。雖然在博士班的時候，做的題目是「中文語音合成系統」，但在畢業後不久，便到聲碩科技股份有限公司工作，主要的產品「說亦通」是一個語音輸入軟體。當時的「說亦通」其實是台大李琳山教授團隊所研發出來的「金聲三號」的一個商業化的版本，運算速度相對較快，需要的記憶體較少，用來訓練的語料庫更多，使用介面更方便等等。這樣的軟體，在當時算是一個全新的商品，市場上也只有 IBM 的 ViaVoice 是主要的競爭對手。由於中文的鍵盤輸入比較複雜，不管是使用注音、倉頡或是其他各種的輸入法，都需要有長時間的練習，才能有較快的速度。因此，中文語音輸入法在當是被視為是一個殺手級的應用，大家十分看好這樣商品的發展，在市場上也引起了一些注意。為了推廣「說亦通」，聲碩到各個資訊展覽的會場去擺攤子，同時訓練了一些業務人員，進行示範，很多人看了示範，覺得非常神奇，也十分實用，就當場買了回去。但是「說亦通」的銷售量並沒有因此逐漸上升，調查一下原因，買回去使用的人，大部分覺得辨識率不好，跟展覽會場看到的差很多，有些人甚至有受騙、上當的感覺。想想也對，以那些業務人員來說，他們也是要經過一段時間的訓練，才能有穩定的辨識率，一般人可能連麥克風都沒戴好，說的時候，更不會注意到要保持適當的速度，如何能有我們在實驗室控制環境下所測試出來的數據呢？為了解決這個問題，聲碩提供了一個小時到府訓練的課程，會使用這樣訓練課程的，大部分是不太懂電腦的人，年紀也偏大，所以還是有很多人即使經過訓練、調整，辨識效果還是很差。而通常如果你希望使用者調整一下說話的速度，或者是稍微注意一下發音，得到的反應總是「我說話就是ㄉㄛ、ㄍㄛ、樣子，是你們的系統太爛了。」這時候你還能說什麼呢？承認吧！現階段語音辨識所使用的語料與語言模型，是沒有辦法解決這樣的問題的。

後來在飛利浦語音科技工作的時候，有一個專案是要做一個手機上面的數字撥號程式，首先做的語言是英語，語音模型是德國總部已經訓練好的，我們只是將應用程式做到實際的手機上，然後進行測試。本來覺得很簡單的一個應用，但測試的時候卻發現數字 6(six)常常講不進去，換個人來測試也一樣，後來只好找個老美來試試看，結果是完全沒有問題。仔細聽聽老美的發音，似乎比較接近 sex 而不像我們平常念的 six。結果大家就嘗試把 6 當成 sex 來念，反而成功的機率變大了，這樣的發現，讓我對自己的英語發音產生了徹底的懷疑，難道以前都念錯了。six 中的母音不就是一個短音的「一」嗎？為什麼會念得有點像「ㄛ」呢？查了一下英語發音的書，原來[i]跟[ɪ]除了長短的差別外，他的發音位置也是不同的，[ɪ]的舌頭位置要稍微向後一點，反而有點接近中文的「ㄛ」，所以念 sex 對的機率才會比較高，有了這樣的認識，再經過不斷反覆的調整與練習，終於找到了那個介於國語「一」跟「ㄛ」中間的一個位置，最後錯誤就很少再發生了。所以，我必須承認，就英語來說，是我念得不好，而不是語音辨識辨認錯誤。

從上面這兩個例子中，我們可以看到技術與應用的極大落差，在實驗室正確率高達 90%以上的系統，在實際應用的時候，可能有五成以上的使用者達不到五成以上的正確率。由於在語音辨識的應用上，電腦必須去適應各種不同類型的使用者，這是一件非常困難的事，而即使是使用者本身的發音錯誤，還是要歸類成辨識錯誤，因此要讓多數使用者滿意是相當困難的。但是如果將語音辨識的技術用在語言學習上，這時候我們便有一個所謂標準的模型，同時期望使用者將他的發音盡量向標準模型靠近，在這樣的狀況下，當使用者是在學習另外一種語言的時候，發音的變異便是一種錯誤，使用者必須使他的發音盡量和標準的發音一致，所以就應用本身而言，電腦可以控制的部分就相對增加了，使用者對整個應用的滿意度也相對可以更高。因此，在離開工業界

到學校服務後，個人便將語音辨識在語言學習上的應用，設定為研究的一個主要方向。

## 2. 過去的發展

近數十年來，第二語言的教學方法主要是由以互動理論（Interactionist Theories）為基礎的溝通式教學法（Communicative Language Teaching）所主導。可是在國內的英語教育環境中，除了私人的英語補習班採用小班的教學外，一般的學校，由於每班的學生眾多，所以很難採用雙向互動的教學方法，因為如果要針對學生個別的問題進行糾正，需要甚多的時間，因此學生普遍發音不甚標準，會話能力也不佳。這幾年英語教育也被安排在國小的課程中，讓學生自小即能開始英語的學習和訓練，然而因為老師本身的發音也未必正確，自然沒有辦法教出發音正確的學生了。我的太太現在就是國小的英語教師，她跟我說過一件糗事，當她在美國唸書的時候，跟外國同學吃飯，吃飽的時候，她就說我吃飽了（I'm full.），結果他們都不懂她為什麼說自己是傻瓜（fool），但其實這也是另外一個我們以為是發音長短的不同，但實際上發音位置卻有蠻大不同的例子。後來她也是花了蠻多的時間，才慢慢把這些問題更正過來。要解決這樣的問題，電腦輔助語言學習系統（Computer Assisted Language Learning System, CALL）似乎可以提供解決的方案，尤其是利用語音辨識（ASR）技術的電腦輔助發音訓練系統（Computer Assisted Pronunciation Training, CAPT）不但可以提供一個沒有壓力的環境，讓學生反覆的練習，同時也能針對學生個別的發音問題，提供回饋與糾正的功能，這無疑是可讓學生的學習效率大幅提昇的方法。

由於 CAPT 的許多好處，國內外的許多學術單位或者商業軟體的出版商，都投入了相當的力量，也完成了許多的系統與商業性的教學軟體，然而在語言教育的學術研究中，卻沒有得到太好的評價。如果根據[1]的意見，主要的問題有兩個：

- (1) 語音辨識的結果沒有達到令人滿意的效果。
- (2) 對於學習者發音的錯誤，沒有辦法正確辨識，並提供適當的回饋。

對語音辨識的辨識率的問題，有相當多的商用教育軟體是採用市場上現有的語音辨識引擎加工而成，可是由於這些引擎本來都是針對母語的使用者，做語音輸入的應用而設計的，所以如果是針對正在學習語言的使用者來說，其辨識率便有相當多的下降，而無法達到原來設計的需求，實際的數據可以參考[2]中的研究。然而這並非表示語音辨識的技術無法使用在語言學習上，只是在使用的的方法上必須經過適當的調整，如果語言學習者的發音，電腦可以很容易就辨識成功，那表示他的發音已經十分標準了，根本不用再學習了。所以我們應該將重點放在第二個問題上，就是如何正確辨識學習者的發音錯誤，並提供適當回饋。

許多的 CAPT 系統會提供一個及時的回饋，將學習者語音的波形或者頻譜顯示在螢幕上，讓使用者自己跟一個所謂標準的發音進行比較，這樣的回饋除了讓使用者覺得很「炫」外，其實際的效果是非常令人質疑的。因為即使是個語音學的專家，也沒有辦法從波形上看出發音是否正確，看頻譜也許可以，但也需要好幾年的訓練，所以對一般的使用者來說，這實在是一點意義也沒有。對於使用了語音辨識技術的 CAPT 系統，通常會提供給使用了一個跟標準模型比較後的評價，通常是一個分數，或者是一個評語，或是一段鼓勵的話。這樣的回饋，根據研究是比較能獲得使用者認同的方式[3]，但是比較具有爭議的是這個分數，跟發音實際的好壞到底有多少關聯。而且，即使這個分數是正確的，但是當使用者念得不對，分數很低的時候，他也不知道要如何修正，所以只好隨機的改變一些方式，看看能不能瞎貓碰上死耗子般的突然變對了，對一些本來就比較發不出來的音，念了一百遍可能還是錯的，所以雖然使用者一開始可能覺得不錯，但矯正的效果也是有限。

就嘗試針對發音的錯誤，提供適當而容易理解的回饋，所進行的研究應該是從一個由歐盟贊助，名為 ISLE 的研究計畫[4][5]開始，這個研究主要針對德國及義大利的英語學習者，針對使用者在發音上的錯誤，提供一個矯正性的回饋，告訴使用者發生錯誤的音，以及修正的方式。而為了要增加語音辨識系統能正確辨識發音錯誤的比率，就必須把可能的錯誤限制在一個較小的範圍

內，因此系統便根據使用者使用的母語，去預測可能發音的錯誤，然後決定辨識比對的範圍。這樣的回饋設計基本上是相當不錯的想法，但是這個系統在測試的時候，似乎辨識錯誤的比率太高，因此給使用者的感覺是「令人沮喪的錯誤回饋，比有用的回饋多」[5]，不過這似乎是一個正確的方向，畢竟根據[6]的研究，在一群外語學習者，參加了一個 120 小時的發音訓練課程後，也只有 50%的學員，有得到一點點的改善，所以發音矯正的訓練不是那麼容易的，如果要設計一個這樣的系統，就必須對一些學習第二語言（中介語）的相關理論，有深入的瞭解。

### 3. 中介語習得相關理論

在早期語言學習的諸多理論中，Lado 的對比分析假說（Contrastive Analysis Hypothesis, CAH）一直都是語言學家研究的重點，不管他是要支持或者反對這個假說。在這個假說中，認為我們如果去分析比對學習者的母語（L1），以及他所要學習的語言（L2）中的同異之處，我們就可以去預測學習者可能碰到的問題，並設計教材讓學習者克服這些問題[7]。在這樣的理論下，L1 在第二語的習得中，通常扮演負面的角色，也就是說，L1 干擾(Interfere)了 L2 的學習，原因是因為我們在學習 L2 時，常會把 L1 的特色轉移(Transfer)到 L2 中，而所謂第二語的習得，也就是在克服 L1 的影響，慢慢地以 L2 的特色取代在原來 L1 的特色，並逐漸與母語說話者相近的一個緩慢過程。

雖然 50 年代末，Chomsky 對語言心理學的研究，使人們對第二語言習得過程中產生的錯誤有了新的認識，人們發現並認為第二語言學習中的錯誤並不能準確地用對比分析的方法預測、解釋和分析，尤其是在語法和詞彙方面。後來的研究也証明了這種懷疑是不無道理的，因此一種新的理論就應運而生了，這就是中介語理論。在中介語理論的研究中認為，學習者在第二語言習得過程中產生錯誤是多方面的因素影響的結果，包括了主要來自母語的負面干擾，所學的有限的目的語知識的干擾，文化因素的干擾，學習方式以及學習態度等方面的影響，其實簡單來說，就是干擾的成分增多了。但是在語音方面的影響，還是以受其母語的影響成分最多，譬如在華語語音教學中，首先倡導用中介語理論來分析外國學生學習華語所產生的錯誤當推魯健驥[8]，其中主要分析的對象，也是受到母語的干擾而產生的語音錯誤。至於在華人學習英語因為母語的關係所容易發生的錯誤，在 Hide 的文章[9]中也有詳細的討論。如果以台灣人學英語的角度來看，Power [10]的文章也提供了一些不錯的資訊，我們就以台灣人的英語發音所受到母語（國語及台語）的影響，做一些說明。

/l/ 常常被發成/i/，因為國語注音符號的「ㄌ」接近/i/，老師都把/l/教成短的「ㄌ」，而沒有去強調兩者舌頭位置，以及鬆緊的不同。同樣的問題，發生在/ʌ /和/ɑ /（都發成注音符號的「ㄩ」）以及/ɒ /和/u/（都被發成注音符號的「ㄨ」），這兩組母音上面，這就是前面所提到「six」跟「full」的問題。「six」的發音對溝通比較沒有影響，因為沒有近似音的字，人的語言模型會自動把他更正回來，但是把「full」當成「fool」的機率就蠻大的。另外/ɛ/, /eI/跟/æ/也常常被發成注音符號的「ㄝ」，所以如果你在廚房要找個鍋子(pan)，別人可能會遞給你一枝筆(pen)。至於在子音方面，/θ /常常會被發成/s/，主要是那個把舌頭伸到牙縫間的動作沒有變成習慣，至於/r/跟/l/的混淆，通常發生在台灣國語的學習者，以我自己為例，以前就常常說「今天天氣很樂」而被朋友糾正。這個問題，在做了許多年國語語音辨識的研究後，已經很少再犯這樣的錯誤了，所以說語音辨識對矯正發音應該有不錯的功效。類似上面這種錯誤類型的整理，還有很多的例子，可以參考 Hide[9]的論文。

根據上面的說明，我們大致瞭解，一個人從小學習一種語言，就會發展出針對這套語言的感知(speech perception)和產生(speech production)的系統。在感知系統方面，會把落在某一個特定範圍內的聲音，歸類為同一個音（嚴格定義稱為音素），在產生系統方面，則會為每一個音發展出一套固定的發音方式，並形成一種習慣。當人們在學習另一種語言時，原來這套語音感知和產生的系統，便會扮演輔助或者干擾的角色，而當它扮演的是干擾的角色的時候，就會產生特定類型的發音錯誤，這些錯誤是可以預期，也可以解釋的，如果我們能夠善用這些知識，就可以減輕語音辨識系統的負擔，讓語音辨識系統達到更好的辨識效果。

#### 4. 語音辨識在 CAPT 中所應扮演的角色

所以語音辨識系統在一個 CAPT 系統中所扮演的角色，應該是去找出使用者發音的錯誤，同時判斷發音錯誤的類型，最後再提供適當的回饋，提供給使用者做修改的依據。然而一般發音錯誤的狀況，可能發生在兩個不同的階段，第一個階段的錯誤是發生在將英文字母轉換成音標的時候，第二個階段則是在發聲器官實際發聲的時候，所產生的錯誤。由於這兩種錯誤的型態不同，產生的原因也不同，必須用不同的方式來處理。在字轉音的錯誤方面，雖然現在國小英語教學，已經逐漸採用自然發音法，但是很多學生對發音還是有很大的障礙，因為即使瞭解了自然發音法，在英文字轉音的規則中，還是有許多的例外。因此這一方面錯誤的型態，並不是十分容易歸納，所以最保險的方法，就是將所有可能的字轉音的規則通通放進來，而不要考慮其前後字母的影響，這樣子比較可能偵測出使用者所有可能的錯誤。至於第二種實際發音時所產生的錯誤，主要是針對學生雖然知道要發哪個音，但是卻發不正確。這部分的錯誤，主要就是前面所說受到母語的影響，通常是來自學生的母語中沒有的音，或者是跟其母語中某個音類似，但有些許差異的音。這些常見的錯誤，在許多第二語言教學的相關文獻中，是很容易找到的，因此如果將這些常常發生的錯誤類型，跟正確的發音，一起放入語音辨識的範圍內，讓語音辨識系統來決定使用者的發音是否正確，假設以前面的 full 的例子來講，讓語音辨識系統去辨認使用者是念 full 還是 fool，這樣的任務，對目前的語音辨識系統來說，是有相當程度的正確比率的[11]。

之前所提到的 ISLE 的研究計畫，最後的研究報告提到語音辨識系統所產生的錯誤回饋，比正確的回饋還要多的例子，或許有人覺得這是語音辨識的辨識率不好的問題，但個人覺得是課程設計的目的不清楚，讓使用者產生的感覺所佔的成分較多。因為它的設計，是讓使用者去念一篇文章，然後，只要辨識系統辨識出使用者的發音有問題，就會產生回饋。在這樣的練習當中，可能發生的錯誤類型就相當的廣泛，因此不斷的產生各種不同的回饋，對使用者來說就會形成一種干擾。尤其是他把我們之前提到兩種不同原因的錯誤類型都混在一起了，這樣不但語音辨識部分的負擔較大，回饋的部分也會變得較為複雜，使用者的注意力就分散了，所以效果並不是很好。因此，如果要解決這些問題，我們就必須設計任務較為單純的課程，讓語音辨識擔任更單純的角色，這些都需要語音辨識程式的開發者，和第二語言教學的老師，更密切的合作，才能達成，尤其是在下個部分所要提到回饋的設計的部分。

#### 5. 發音錯誤的回饋設計

如果我們希望製作一個程式，能夠像一個老師一樣來幫助學生改善它的發音，語音辨識的部分，只能勉強算是這位老師的耳朵，但這其實是蠻重要的，因為老師可以用一張嘴，同時教三十個學生，但卻不能用兩個耳朵，去分辨三十個學生的發音是否標準。因此利用語音辨識的技術，來輔助發音的訓練，其實是蠻能補足目前英語教學的不足。只是在聽出學生的問題後，要用什麼樣的方式，來讓學生瞭解自己的錯誤，並進行改正，因為即使是一個老師，也是需要相當的教學經驗才能完成這個任務。如果只是把發音的方式，不斷的重複教導、示範，讓學生不斷的嘗試，或許學生可以用亂槍打鳥的方式，發出幾次正確的發音，但可能很快就忘記了。所以通常一個有經驗的老師，會有許多教學的技巧，根據不同的錯誤方式，提供不同的指示，讓學生可以比較快的抓到訣竅，並且記住發音的方式，之後才能不斷的練習，並形成習慣。

這樣的發音糾正的方式，我們假設以砲兵發射大砲來當作例子，能夠第一發就集中目標的機率是很少的，但是如果能夠根據第一發的彈著點進行修正，這樣誤差就會有系統的減少，很快就能命中目標了。同樣的道理，對應到發音的練習上，學生通常第一發打的位置，就是母語中較相近的那個音，如果語音辨識可以辨識出這個現象，那麼他的發音跟實際目標的誤差其實系統就已經知道了，剩下的任務，只是如何讓學生移動他的發音器官到正確的位置。這部分其實是整個教學的重點，但也是目前大部分的 CAPT 系統中，所最缺乏的。問題在哪裡呢？大部分的 CAPT 系統，並沒有去設定學生的對象，所以所有的學生，不管他的母語背景如何，基本上共用同一套的回饋系統，這樣的效果一定比不上，善用他原來母語的感知和產生系統的回饋。舉個例子：台灣的學

生常常把/æ/發成/ /, 如果單純叫學生把嘴巴張大點, 舌頭後退一點, 那麼學生通常會在原地打轉了很久, 因為這個「一點」是多大, 並沒有定義, 學生通常只會在他習慣的位置打轉。如果你告訴學生, 其實有點像「Y」, 這時候移動速度就會快了, 即使過頭了, 要修正回來也是比較容易。所以, 回饋的設計, 一定要根據學生原來的母語來設計, 才比較容易達到效果。

## 6. 英語發音訓練系統之雛形

從上述的說明當中, 我希望傳達的概念是語音辨識的技術, 是可以幫助我們學習另外一種語言, 但是成效的好壞, 決定在如何的設計。但是如何達到一個好的設計, 其實需要跨領域的合作, 與不斷的研究與嘗試。我把個人上述的想法, 做了一點嘗試, 完成一個英語發音訓練系統-「牙牙學語機」, 詳細的內容可以參考[12]中的說明。當初之所以將這個系統, 做在一個手持式的設備上, 主要是希望使用者可以隨時來練習, 或者將這樣的功能, 加入到現在十分普及的電子字典中, 這樣相對於在 PC 執行的系統來說, 使用起來會更加的方便。另外, 透過網路也是一個十分方便, 且容易推廣的方式, 因此這部分我們也做了一個在網路上展示的系統, 可以讓有興趣的使用者, 上去使用看看。由於這只是初步的嘗試, 還希望各位會員能多多提供一些意見, 作為改進的參考, 或者有興趣者, 一起來進行相關議題的研究。這個網站的網址是:

<http://mcu.edu.tw/~fuchiang/DemoEng/default.htm>

使用時, 請務必參考使用說明, 有問題, 請不吝跟我聯絡。

## 參考資料

1. T. M. Derwing, M. J. Munro and M. Carbonaro, "Does Popular Speech Recognition Software Work with ESL Speech?", TESOL Quarterly 34, pp. 592-603, 2000
2. D. Coniam, "Voice Recognition Software Accuracy with Second Language Speakers of English", System 27, pp. 49-64, 1999
3. ISLE Deliverable 1.4, Pronunciation training: requirements and solutions, 1999, <http://nats-www.informatik.uni-hamburg.de/~isle/public/D14/D14.html>
4. Interactive Spoken Language Education (ISLE), A Language Engineering Project of EU, <http://nats-www.informatik.uni-hamburg.de/~isle/index.html>
5. Menzel, W., Herron, D., Bonaventura, P., Morton, R., "Automatic detection and correction of non-native English pronunciations", Proceedings of InSTILL 2000, pp. 49-56
6. Acton, W. "Changing Fossilized Pronunciation", TESOL Quarterly 18, pp. 71-85, 1984
7. R. Lado, "Linguistics Across Cultures", University of Michigan Press, 1957
8. 魯健驥, "中介語理論與外國人學習漢語的語音偏誤分析", 語言教學與研究, 第三期, pp. 44-56, 1984
9. Hide, O., "Interlanguage phonology: implications for a remedial pronunciation course for Chinese learners of English", Antwerp papers in linguistics, Vol. 100, pp. 17-46, 2002
10. Power, T., "English Pronunciation for Taiwanese Speaker", <http://www.btinternet.com/~ted.power/11taiwanese.html>
11. Fu-chiang Chou, "On-line English Pronunciation Training System with Automatic Speech Recognition Technologies", Proceeding of ED-MEDIA, pp. 3819-3824, 2005
12. Fu-chiang Chou, "Ya-Ya Language Box - A Portable Device for English Pronunciation Training with Speech Recognition Technologies", Interspeech, pp. 169-172, 2005