中華民國計算語言學學會
通訊 ACLCLP
發行人：吳宗憲
主　編：王新民
執行編輯：黃　琪
The Association for Computational Linguistics and Chinese Language Processing 第二十一卷第二期
E-Mail：aclclp@hp.iis.sinica.edu.tw
地 址：台北市研究院路二段128號中研院資訊所
Website : http://www.aclclp.org.tw
電 話：(02)2788-3799 ext:1502
傳 眞：(02)2788-1638
劃 撥：19166251

本 期 要 目

壹、**ROCLING 2010 CALL FOR PAPERS** 第二頁
貳、**2010** 語音訊號處理研討會議程及報名表 第三~四頁
參、**ISCSLP 2010 CALL FOR PAPERS** 第五頁
肆、專文- 使用反射係數之頻譜包絡內插方法(古鴻炎、蔡哲彰) 第六~十六頁

IJCLCLP & ROCLING have been included in ACL Anthology

Thanks to the effort of former ACLCLP President, Prof. Hsin-Hsi Chen, and ACL Anthology Editor, Prof. Min-Yen Kan, the papers published by the IJCLCLP journal and the annual ROCLING conference of ACLCLP have now been listed in the ACL Anthology, a digital archive of research papers in computational linguistics. The ACL Anthology (http://www.aclweb.org/anthology/) currently hosts over 17,600 papers and is the most complete archive in this field. With a single search interface powered by Google, the papers (in Chinese or English) from IJCLCLP and ROCLING can now be accessed with other related papers in the broad field of computational linguistics, a valuable benefit to both the readers and authors.

This archive hosts the IJCLCLP papers from 1996 to 2009 and ROCLING proceedings from 1988 to 2009, a total of more than 600 papers highly related to Chinese language and speech processing. This collection will be updated regularly in the ACL Anthology, once new papers have been published by the journal and the conference.

Currently, ROCLING and the IJCLCLP journal are listed together on the same page. This will be changing in a few months when the new ACL Anthology interface is in place.

New book: Computer Processing of Asian Spoken Languages

Computer Processing of Asian Spoken Languages Edited by Prof Shuichi Itahashi and Prof Chiu-yu Tseng, this book contains recent research papers by 120 speech researchers in Asia. It makes possible to overview the trends of speech processing research in Asia. This book serves as a tribute to how the problems of computer processing of the languages spoken in Asia were overcome and as an inspiration for speech technology development in the future. It is a one-of-a-kind reference resource for the entire speech community as Asia continues to play a much more active role in the international arena.

The book will be available at amazon.com soon. For more details, please visit: http://www.sig-cslp.org/documents/O-C book publicity leaflet.pdf.

ROCLING-2010

「第二十二屆自然語言與語音處理研討會」由國立暨南國際大學資訊工程學系、電機工程學系、及本會聯合主辦，謹訂於九月一日(週三)～二日(週四)假南投縣埔里鎮國立暨南國際大學科技學院第一演講廳舉行。ROCLING 會議論文已被 ACL Anthology 收錄，歡迎大家踴躍投稿，徵稿啓事請參閱本刊第二頁。

Conference on Computational Linguistics and Speech Processing

第二十二屆自然語言與語音處理研討會

September 1-2, 2010, National Chi Nan University, Nantou, Taiwan

http://nlp.csie.ncnu.edu.tw/rocling-2010/

CALL FOR PAPERS

Conference Chair

Jing-Shin Chang
National Chi Nan University

Program Committee

Shih-Hung Wu, Co-Chair
Chaoyang University of Technology

Jeih-Weih Hung, Co-Chair
National Chi Nan University

Guo-Wei Bian
Huafan University

Chia-Hui Chang
National Central University

Jason S. Chang
National Tsing Hua University

Berlin Chen
National Taiwan Normal University

Chia-Ping Chen
National Sun Yat-Sen University

Hsin-Hsi Chen
National Taiwan University

Keh-Jiann Chen
Academia Sinica

Kuang-Hua Chen
National Taiwan University

Sin-Horng Chen
National Chiao Tung University

Jen-Tzung Chien
National Cheng Kung University

Hung-Yan Gu
National Taiwan University of Science and Technology

Wen-Lian Hsu
Academia Sinica

Jyh-Shing Jang
National Tsing Hua University

Chih-Chung Kuo
Industrial Technology Research Institute

June-Jei Kuo
National Chung Hsing University

Chao-Lin Liu
National Chengchi University

Jyi-Shane Liu
National Chengchi University

Ren-Yuan Lyu
Chang Gung University

Tzong-Han Richard Tsai
Yuan Ze University

Yuen-Hsien Tseng
National Taiwan Normal University

Hsiao-Chuan Wang
National Tsing Hua University

Hsin-Min Wang
Academia Sinica

Yih-Ru Wang
National Chiao Tung University

Chung-Hsien Wu
National Cheng Kung University

Gin-Der Wu
National Chi Nan University

Ming-Shing Yu
National Chung Hsing University

The 22nd ROCLING Conference will be held at National Chi Nan University, Nantou, on September 1-2, 2010. Sponsored by Association for Computational Linguistics and Chinese Language Processing (ACLCLP), ROCLING is the most historied and major conference in the broad field of computational linguistics, speech processing, and related areas in Taiwan.

ROCLING XXII will be co-hosted by the Department of Computer Science and Information Engineering and the Department of Electrical Engineering, National Chi Nan University. The two-day conference will feature invited talks, paper, and poster sessions.

ROCLING XXII invites submissions of original and unpublished research papers on all areas of computational linguistics, natural language processing, and speech processing, including, but not limited to, the following topic areas.

Topics of Interest:

- cognitive/psychological linguistics
- discourse/dialogue modeling
- information extraction/text mining
- information retrieval
- language understanding/generation
- lexicon/morphology
- machine translation/multilingual processing
- named entity recognition
- NLP applications/tools/resources
- phonetics/phonology
- question answering
- semantic web
- semantics/pragmatics
- speech analysis/synthesis
- speech recognition/understanding
- spoken dialog systems
- spoken language processing
- syntax/parsing
- text/speech summarization
- web knowledge discovery
- word segmentation/POS tagging
- others

Submission Guidelines:

Prospective authors are invited to submit full papers of no more than 25 A4- sized pages in PDF format (please go to the conference website to download the paper guideline and template). Papers will be accepted only by electronic submission through the conference website. The submitted papers should be written in either Chinese or English, and in single column, single-spaced format. The first page of the submitted paper should bear the items of paper title, author name, affiliation, and email address. All these items should be properly centered on the top, followed by a concise abstract of the paper.

Papers should be made in PDF format and submitted online at:
http:// (to be decided)

Best Paper Award:

The best paper will be selected and announced at ROCLING XXII.

Important Dates:

Preliminary paper submission due:	**July 2, 2010**
Notification of acceptance:	August 6, 2010
Final paper due:	August 15, 2010
Conference date:	September 1-2, 2010

Sponsors:

Association for Computational Linguistics and Chinese Language Processing (ACLCLP)
Department of Computer Science and Information Engineering,
National Chi Nan University
Department of Electrical Engineering, National Chi Nan University

2010 語音訊號處理研討會
Advanced Speech Enabling Human-Machine Interface

會議時間：99 年 5 月 4 日 (二)
會議地點：嘉義大學蘭潭校區資工館國際會議廳

語音訊號處理研討會是中華民國計算語言學學會定期舉辦的學術交流盛會，今年度的會議將以前瞻的語音人機介面(Advanced speech enabling human-machine interface)爲主題，邀請國內外知名的學者專家介紹此領域最新的研發成果與學術界及工程界分享。所邀請之演講者包括來自香港中文大學的蒙美玲教授(Prof. Dr. Helen M. Meng)，台灣師範大學陳柏琳教授，台北科技大學廖元甫教授，長庚大學呂仁園教授，與清華大學張智星教授。講授語音訊號處理、語音檢索技術、麥克風陣列與語音處理、台語語音辨認以及語音評估等主題。除了以上演講外，本次會議同時舉辦國科會成果發表與相關系統展示，對於語音相關工作者有極爲豐富之內容。

99 年 5 月 4 日		
時　　間	場 次	活動內容
9:00-9:30	報　　到	
9:30-9:40	開幕致詞	主講人：柯建全 教授 (嘉義大學 理工學院院長) 吳宗憲 教授 (成功大學 電資學院副院長)
9:40-10:50	第一場	講　題：Advanced Speech Signal Processing (Tentative) 主講人：Helen M. Meng 教授 (香港中文大學 工程學院副院長)
Coffee Break		
11:10-12:20	第二場	講　題：語音評分的方法與應用 主講人：張智星 教授 (清華大學 資工系)
12:20-13:20	Lunch Time	
13:20-14:40	第三場	講　題：現階段語音檢索與摘要研究之介紹 主講人：陳柏琳 教授 (臺灣師範大學 資工系)
14:40-16:00	第四場	講　題：麥克風陣列與語音處理 主講人：廖元甫 教授 (台北科技大學 電通系)
Coffee Break		
16:20-17:40	第五場	講　題：台語語音辨識 主講人：呂仁園 教授 (長庚大學 資工系)
閉幕		

2010 語音訊號處理研討會

報名表

會議時間：2010 年 5 月 4 日(星期二) 9:00~18:00
會議地點：嘉義大學蘭潭校區資工館國際會議廳
主辦單位：嘉義大學資訊工程系、中華民國計算語言學學會
協辦單位：國科會工程處工程科技推展中心

姓　　名		單　　位	
電　　話		E-mail	
聯絡地址			
收據抬頭			
報　名　費	一般人士：　□會員 500 元　□非會員 600 元 學生：　□會員 300 元　□非會員 500 元 (報名費包含講義、午餐及茶點)		

1. 學生非會員請附學生證影本
2. 報名截止日 **4/26** 日星期一（現場報名加收 100 元）
3. 繳費方式：

   □　郵政劃撥：劃撥帳號 19166251、戶名：中華民國計算語言學學會

   □　信用卡：Card Type : (Visa / Master-card / JCB)

   信用卡持有人：_____發卡銀行：_____

   卡號：_____-_____-_____-_____

   有效期：_____（月/年）　卡片後三碼：_____

   持卡人簽名：_____（簽名方式請與信用卡相同）

   □　午餐請備素食

4. 報名表請傳真或 E-mail 至：

   傳真：(02) 2788-1638、E-mail: aclclp@hp.iis.sinica.edu.tw
   電話：(02) 2788-3799 分機1502，連絡人：黃琪 小姐
   (大會網址：http://office.csie.ncyu.edu.tw/sws2010/)

ISCSLP 2010

Nov. 29 - Dec. 3, 2010 Tainan & Sun Moon Lake, Taiwan

Honorary Chair
Michael Lai, *Natl. Cheng Kung Univ., Tainan*

General Co-Chairs
Jhing-Fa Wang, *Natl. Cheng Kung Univ., Tainan*
Chiu-Yu Tseng, *Academia Sinica, Taipei*

Steering Committee Chair
Chin-Hui Lee, *Georgia Institute of Tech., USA*

International Advisory Committee
Sin-Horng Chen, *Chiao Tung Univ., Hsinchu*
P. C. Ching, *The Chinese Univ. of Hong Kong, H.K.*
Hsiao-Wen Hon, *Microsoft Research Asia, Beijing*
Xuedong Huang, *Microsoft Corp., Redmond*
Biing Hwang Juang, *Georgia Institute of Tech., USA*
Chin-Hui Lee, *Georgia Institute of Tech., USA*
Lin-Shan Lee, *Natl. Taiwan Univ., Taipei*
Mao-Can Lin, *Chinese Academy of Social Sciences, Beijing*
Hsiao-Chuan Wang, *Natl. Tsinghua Univ., Hsinchu*
Ren-Hua Wang, *Univ. of Science & Tech. of China, Hefei*
Victor Zue, *MIT, USA*
Frank Soong, *Microsoft Research Asia, Beijing*
Lee-Feng Chien, *Google, Taipei*
Satoshi Nakamura, *NICT, Japan*
Hiroya Fujisaki, *Univ. of Tokyo, Japan*
Sadaoki Furui, *Tokyo Institute of Tech., Japan*

Special Consultants
Chung-Hsien Wu, *Natl. Cheng Kung Univ., Tainan*
Qiang Huo, *Microsoft Research Asia, Beijing*
Helen Meng, *The Chinese Univ. of Hong Kong, H.K.*
Haizhou Li, *Institute for Infocomm Research, Singapore*
Yu Hu, *Univ. of Science & Tech. of China, Hefei*
Frank Soong, *Microsoft Research Asia, Beijing*

General Secretary
Shi-Huang Chen, *Shu-Te Univ., Kaohsiung*

Technical Program Co-Chairs
Hsin-Min Wang, *Academia Sinica, Taipei*
Brian Kan-Wing Mak, *The H. K. Univ. of Sci. and Tech., H. K.*

Plenary Chair
Chung-Hsien Wu, *Natl. Cheng Kung Univ., Tainan*

Special Session Co-Chairs
Fang Zheng, *Tsinghua Univ., Beijing*
Tan Lee, *The Chinese Univ. of Hong Kong, H.K.*
Berlin Chen, *Natl. Taiwan Normal Univ., Taipei*

Tutorial Co-Chairs
Jianhua Tao, *Chinese Academy of Sciences, Beijing*
Chia-Ping Chen, *Natl. Sun Yat-sen Univ., Kaohsiung*

Exhibition Co-Chairs
Carl Kuo, *Info. & Com. Research Laboratories, ITRI, Hsinchu*
Lynn Shen, *Delta Electronics Inc., Taipei*

Publicity Co-Chairs
Yuan-Fu Liao, *Natl. Taipei Univ. of Tech., Taipei*
ZhongHua Fu, *North Western Polytechnical Univ., Xi'an*
Lei Xie, *North Western Polytechnical Univ., Xi'an*

Publication Co-Chairs
Jui-Feng Yeh, *Natl. Chiayi Univ., Chiayi*
Bin Ma, *Institute for Infocomm Research, Singapore*

Local Arrangement
Vivian Hou, *Natl. Cheng Kung Univ., Tainan*
David Fang, *Natl. Cheng Kung Univ., Tainan*
Po-Chuan Lin, *Tung Fang Institute of Tech., Kaohsiung*
Jia-Ching Wang, *Natl. Central Univ., Taoyuan*
Hsien-Chang Wang, *Chang Jung Christian Univ., Tainan*
Shun-Chieh Lin, *Industrial Tech. Research Institute, Tainan*

Conference Administration
Center for Industry Elite Training and Convention Management, Natl. Cheng Kung Univ., Tainan

Information
Vivian Hou ,*Tel:+886-6-2096455*
Fax:+886-6-2381422
service@iscslp2010.ncku.edu.tw

Call For Papers

ISCSLP is the flagship conference of ISCA SIG-CSLP (International Speech Communication Association, Special Interest Group on Chinese Spoken Language Processing). ISCSLP2010 will be held during November 29 - December 3, 2010 in Tainan and Sun Moon Lake, Taiwan hosted by National Cheng Kung University.

Tainan, located in south-western Taiwan, is the city of cultural origin. There are many historical places and heritage sites. In addition, Tainan is a modern city with various shopping centers, department stores, and night markets. It will be a wonderful opportunity to experience Taiwanese cultures when you visit Tainan. **Sun Moon Lake**, the largest lake located in central Taiwan, is a beautiful alpine lake, with its eastern part rounded like the sun and the western side shaped like a crescent moon. Its crystalline, emerald green body of water reflects the hills and mountains surrounding on all sides. Its natural beauty is further enhanced by numerous cultural and historical sites.

We invite your participation in this premier conference, where the language from ancient civilizations embraces modern computing technology. ISCSLP 2010 will feature world-renowned plenary speakers, tutorials, exhibits, and a number of lecture and poster sessions on the following topics:

- Speech Production and Perception
- Phonetics and Phonology
- Speech Analysis
- Speech Coding
- Speech Enhancement
- Speech Recognition
- Speech Synthesis
- Language Modeling and Spoken Language Understanding
- Spoken Dialog Systems
- Spoken Language Translation
- Speaker and Language Recognition
- Indexing, Retrieval and Authoring of Spoken Documents
- Multi-Modal Interfaces
- Spoken Language Resources and Technology
- Applications of Spoken Language Processing Technology

Official Language & Publication

- The official language of ISCSLP is English.
- All papers accepted will be included in IEEE Xplore and indexed by EI Compendex

Paper Submission

- Authors are invited to submit original, unpublished work in English.
- Papers should be submitted via http://conf.ncku.edu.tw/iscslp2010/paper.htm.
- Each submission will be reviewed by two or more reviewers.
- At least one author of each paper is required to register.

Schedule

Full paper submission by	July 15, 2010
Notification of acceptance by	Aug. 30, 2010
Camera ready papers by	Sep. 13, 2010
Registration to cover an accepted paper by	Oct. 13, 2010

使用反射係數之頻譜包絡內插方法

A Spectral-envelope Interpolation Method Using Reflection Coefficients

古鴻炎　蔡哲彰

國立台灣科技大學 資訊工程系

1. 前言

這裡”頻譜包絡”指的是振幅頻譜包絡(spectral magnitude envelope)，一個例子如圖 1 裡較爲平滑的曲線，它是由一個/bi/音音框(frame)的 DFT (discrete Fourier transform)頻譜所估計出的。在一些語音處理的子領域中，會需要在兩條頻譜包絡曲線之間作內插，以求得兩者之間過渡的頻譜包絡。例如在使用單元串接之語音合成系統裡，爲了避免在串接點發生頻譜不連續(spectral mismatch)的現象，就需要作頻譜包絡之內插，再依據內插出的頻譜包絡去產生語音信號；雖說過去有一些方法被提出來減輕串接點上頻譜不連續的問題[1]，但那並未根本地解決問題。在歌聲合成的系統裡，由於可能的音高(pitch)、音長(duration)、音節之組合數量更爲龐大，因此當從語料庫(corpus)選取合成單元來作串接時，串接點上幾乎都會發生音高軌跡(pitch contour)不連續、及頻譜的不連續，所以有需要在串接點附近作頻譜包絡內插。此外在語音辨識系統裡，一種可能的語者調適(speaker adaptation)方法是，將某兩語者的音素(phoneme)頻譜包絡作內插，用以逼近目前使用者的音素頻譜包絡。

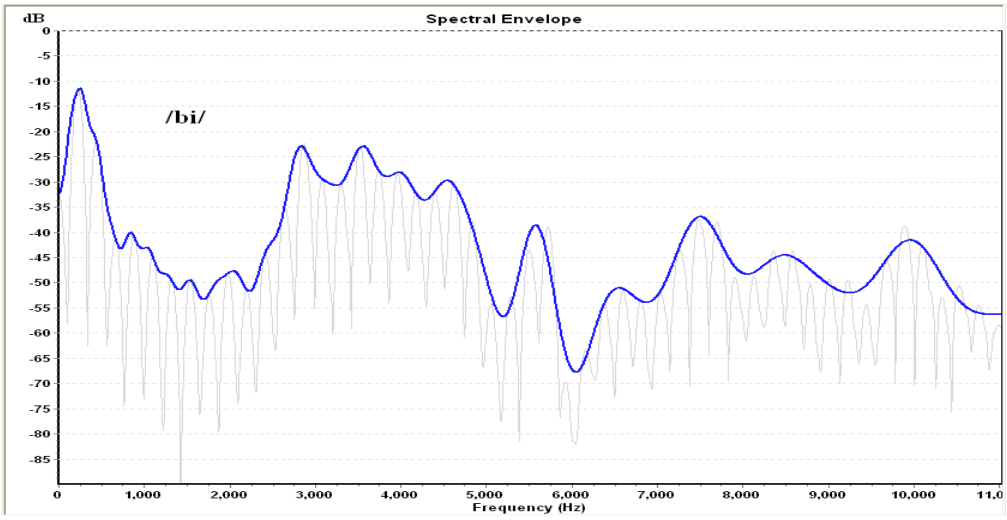


圖 1　/bi/發音的一個音框的頻譜包絡

在作頻譜包絡的內插之前，參與的語音音框的頻譜包絡必需先被估計出來，關於頻譜包絡的估計，過去已有一些方法被提出[2, 3]，例如以 LPC (linear prediction coding)全極(all pole)模型之頻率響應(frequency response)來逼近頻譜包絡[4]；在 DFT 頻譜上以平滑化的倒頻譜(cepstrum)曲線作迭代改進之”True-Envelope”估計法[3, 5]；消除信號本身週期性之干

擾、而具有極高準確性之 STRAIGHT 估計法[6]；以及我們最近嘗試作改進之離散倒頻譜(discrete cepstrum, DCep)估計法[7, 8]。雖然 STRAIGHT 估計法的準確性很高，而 True-Envelope 估計法的準確性也很好，但是兩者需求的計算量很大，目前要以個人電腦的計算能力來達成即時處理，是非常困難的，因此我們仍然決定採用 DCep(離散倒頻譜)之估計法。由於本篇文章的重點在於頻譜包絡之內插，所以頻譜包絡估計方法的細節，就不在這裡介紹了。

2. 內插的方法、領域、與效果

談到內插的方法，一般人直覺上會想到的是線性內插(linear interpolation)，如果使用線性內插就能達成所要求的效果，那麼就不必把事情複雜化。因此對於頻譜包絡內插的問題，首先考慮我們究竟要達成什麼效果？其實從聲響音素學(acoustic phonetics)的知識[9]可知道，當連續發音兩種母音音素時，如連續發/a/與/i/，這兩個音素之間會有共發聲(coarticulation)的現象，而共發聲現象表現在聲譜(spectrogram)上，就是兩音素的共振峰軌跡(formant trajectory)在音素的邊界附近，會以平滑移動的方式相互靠近，一個例子如圖 2 所示，所以頻譜包絡內插要達成的效果就是，當逐漸改變內插比例值時，所得到的一序列頻譜包絡能夠讓共振峰頻率逐漸改變，而形成平滑移動之共振峰軌跡。在確定所需求的效果之後，接著就可考慮執行內插的領域(domain)、及內插的方法。

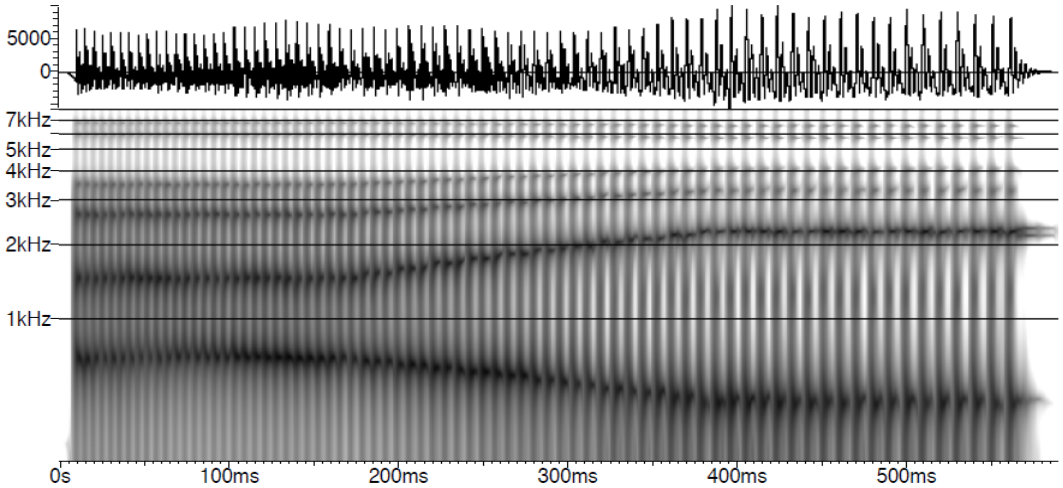


圖 2　共振峰軌跡以平滑移動方式相互靠近[10]

執行內插的領域可以選擇在頻域，即直接使用頻譜包絡，或是選擇在其它頻譜係數的領域，例如：LPC 係數、LSF(line spectrum frequency)係數、反射(reflection)係數，DFT 倒頻譜係數、DCep 係數、…等等，至於選擇那一種係數較好？應採取那一種內插方法來作搭配？則不是立刻可以推想得知。不過，如果我們直接對兩條頻譜包絡曲線作線性內插，則可立刻推知其結果是，共振峰軌跡會以平行方式作淡進淡出(cross fading)，一個例子如圖 3 所示。此外，由前人的論文也已經知道[10]，若選擇 LSF 係數來作線性內插，則會發生某些共振峰軌跡可以平滑地移轉、連接，而某些共振峰軌跡卻會中斷的不一致情形，例子如圖 4 裡經 LSF 係數內插/a/音框與/i/音框所得到的聲譜圖。

過去已經有人研究頻譜包絡內插的方法，並且可以達成前面所說的效果，Pfitzinger 的 DFW (dynamic frequency warping)爲基礎的方法[10]、和 Ezzat 等人的聲訊流(audio flow)爲基礎的方法[11]，都是直接在頻譜包絡上進行內插，但是使用了不同的內插方法。DFW 的觀念是，透過動態頻率校正來建立兩條包絡微分曲線之間的頻率軸對應關係；聲訊流的觀念，則是從 optical flow 的觀念衍生而來，應用時也是在包絡微分曲線上進行聲訊流的計算。如果說還有另外一種頻譜包絡的內插方法，我們覺得基於統計模型、及最大似然(maximum likelihood)原則的方法也算是一種，例如最近常被使用於作語音合成之 HMM (hidden Markov model)模型及其音框頻譜係數的求解方法[12]，但是統計模型之參數需要收集大量的語料來作估計，這是它的缺點。

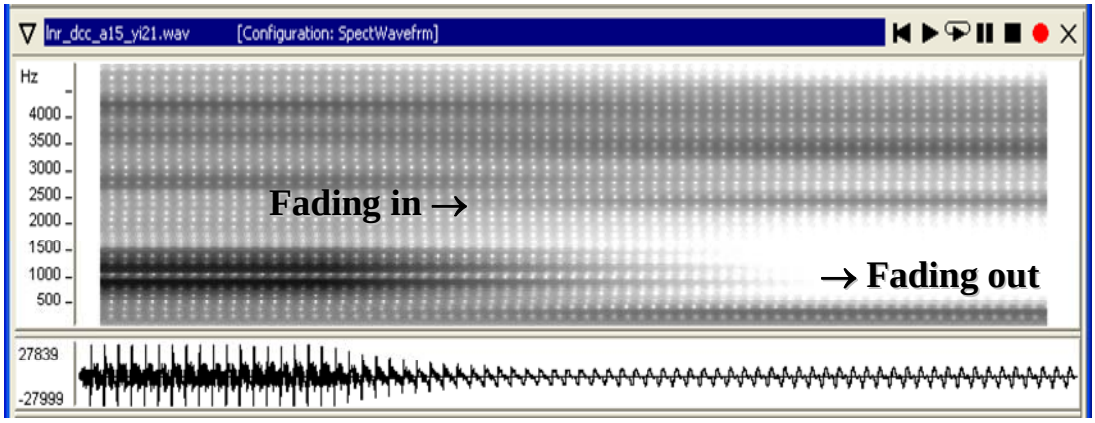


圖 3　共振峰軌跡淡進淡出之例子

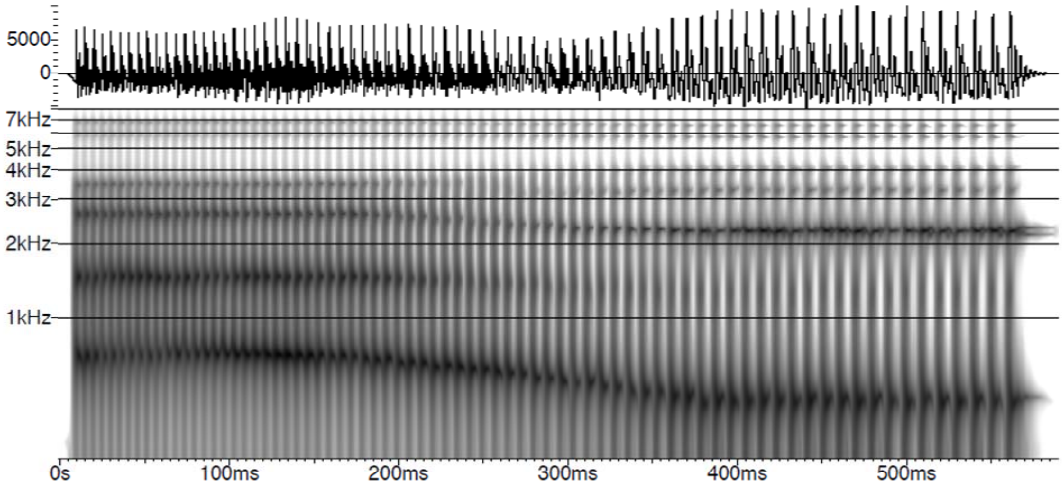


圖 4　共振峰軌跡**平滑連接**與**中斷**並存之聲譜圖[10]

3.　反射係數與頻譜包絡內插

尋找一種恰當的頻譜參數，再對它作簡單的線性內插處理，即可達成前述的頻譜包絡內插效果，一直是我們追求的目標。經過對幾種頻譜係數作實驗驗證，我們最後發現反射係數就是所要找的頻譜係數。反射係數具有的物理意義是，當把聲道(vocal track)模式化(modeling)成多節管(multi-tube)之聲響(acoustic)模型，如圖 5 所畫的，則第 k 個反射係數

C_k的意義就是，第 k 與 k+1 節管子的截面積 B_k與 B_{k+1}的差值與和值的比值，即

$$C_k = \frac{(B_{k+1} - B_k)}{(B_{k+1} + B_k)} \quad . \tag{1}$$

從反射係數所具有的物理意義，我們認爲這可用以解釋，爲什麼對反射係數作簡單的線性內插，就可以達成所想要的頻譜包絡內插之效果。

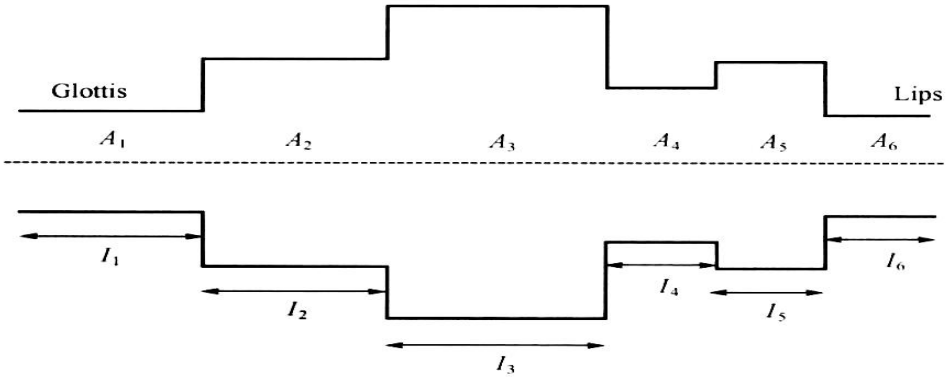


圖 5　聲道之多節管模型[13]

3.1 頻譜包絡至反射係數之轉換

當從一個音框估計得到一條頻譜包絡曲線 S_i, i=0, 1, …, N-1, N 爲 DFT 轉換的點數，本文中設 N=512，接著我們可對 S_i取平方而得到功率頻譜(power spectrum)，然後作 N 點的離散反傅利葉轉換(inverse DFT)，以求得自相關(autocorrelation)係數 R_n [14, 15]，離散反傅利葉轉換的公式爲

$$R_n = \frac{1}{N}\sum_{k=0}^{N-1} |S_k|^2 \cdot e^{j\frac{2\pi}{N}kn} \ , \quad n = 0,1,\cdots,N-1, \tag{2}$$

實作上則是以反快速傅利葉轉換(inverse FFT)來計算公式(2)。依據自相關係數 R_n，接下來可再使用 Levinson-Durbin (LD)演算法[9, 13, 14]來求取 p 階的 LPC 係數 α_k 與反射係數 C_k, k=1, 2, …, p，LD 演算法的詳細步驟如下：

步驟(a)　　令 $E^{(0)} = R_0$,　$m = 1$ (3)

步驟(b)　　$$C_m = \frac{1}{E^{(m-1)}}\left(R_m - \sum_{i=1}^{m-1} \alpha_i^{(m-1)} \cdot R_{m-i}\right) \tag{4}$$

步驟(c)　　$$\alpha_m^{(m)} = C_m \tag{5}$$

步驟(d)　　$$\alpha_i^{(m)} = \alpha_i^{(m-1)} - C_m \cdot \alpha_{m-i}^{(m-1)}, \quad i = 1, ..., m-1 \tag{6}$$

步驟(e)　　$$E^{(m)} = E^{(m-1)}\left(1 - C_m \cdot C_m\right) \tag{7}$$

步驟(f)　　如果 $m < p$，則令 $m = m$+1，再回到步驟(b)

最後令 $C_0 = R_0$，以保留能量資訊。

3.2 反射係數至頻譜包絡之轉換

當以線性內插(或其它內插方式)得到一組新的反射係數 C_i 之後，接著可執行前述的 LD 演算法，不過步驟(b)要略過，因爲反射係數是已知的，如此就可求得 p 階的 LPC 係數 $\alpha_i^{(p)}$，i=1,2, …, p。依據所求得的 LPC 係數，就可用以建立一個全極模型，然後計算此全極模型的頻率響應，就可求得對應的頻譜包絡，也就是藉由全極模型來定義一組反射係數對應的頻率響應，全極模型的正規化(normalized)頻率響應的計算公式爲

$$Q_k = 1 \Big/ \left| 1 - \sum_{n=1}^{p} \alpha_n^{(p)} \cdot e^{-j\frac{2\pi}{N}kn} \right| \quad , \; k = 0,1,\cdots,N-1 \tag{8}$$

理論上只要階數夠高(p 夠大)，頻譜包絡裡的零點(zero)效應仍可被極點(pole)所逼近[9]。

當使用公式(8)計算出頻譜包絡 Q_k 之後，會發現它和真實的頻譜包絡之間有能量上的偏移，因此需要對 Q_k 乘上一個能量調整值 u，$u = (E_{org} / E_{lpc})^{0.5}$，其中 E_{org} 表示原始頻譜包絡的能量(可由 C_0 取得)，E_{lpc} 則表示 LPC 頻譜包絡 Q_k 的能量。

4. 語音信號再合成

令 SA_k, SB_k 表示兩條給定的頻譜包絡曲線，當對它們轉換出的反射係數作內插，我們可得到一序列時間點上的頻譜包絡曲線 $S_k^{(y)}$，k = 0, 1, …, N-1，y = 0, 1, …, 120，在此 y 表示時間軸上的第 y 個控制點或音框。爲了方便說明，這裡先採取線性內插的方式，也就是令

$$CT(S_k^{(y)}) = \begin{cases} CT(SA_k), \; if \;\; 0 \le y < 30 \\ CT(SA_k) \cdot \frac{90-y}{60} + CT(SB_k) \cdot \frac{y-30}{60}, \; if \;\; 30 \le y \le 90 \\ CT(SB_k), \; if \;\; 90 < y \le 120 \end{cases} \tag{9}$$

在公式(9)裡 $CT(\cdot)$表示頻譜包絡至反射係數之轉換。

對於內插出的頻譜包絡序列 $S_k^{(y)}$，我們希望用以合成出語音信號，這樣就可以聽覺感受/a/音音框的 SA_k 與/i/音音框的 SB_k，當合成出語音信號時，是否會有雙母音/ai/的感受，並且合成的語音信號作聲譜分析後，也可用以觀察共振峰軌跡的走勢。因此，我們接著採用 HNM (harmonic-plus-noise model)信號模型[16]，並且依據頻譜包絡來決定各個諧波(harmonic)與雜音(noise)的振幅，然後再據以產生出語音信號。

設 $F_0^{(y)}$是對第 y 個音框所給定的基本頻率，則此音框第 n 個諧波的頻率就是 $n \cdot F_0^{(y)}$，而它的振幅則可依頻譜包絡曲線 $S_k^{(y)}$去決定，一個簡單方法是找出頻譜包絡曲線之頻率軸上最靠近 $n \cdot F_0^{(y)}$的相鄰兩點上的振幅去作線性內插。另外，我們直接設定 MVF (maximum voiced frequency)爲 6,500Hz，並且設定兩音框之間間隔 100 個樣本點，而取樣率則設爲

22,050。爲了方便說明，令第 i 個音框所求出的諧波參數是 f_k^i, a_k^i, k=1, 2, …, L^i, f_k^i 與 a_k^i 分別表示第 k 個諧波的頻率與振幅；再令第 i+1 個音框所求出的諧波參數是 f_k^{i+1}, a_k^{i+1}, k=1, 2, …, L^{i+1}。如此，當要合成第 i 和第 i+1 音框之間時刻 t 的諧波信號之樣本 $h(t)$時，我們先以如下公式作線性內差，

$$
\begin{aligned}
f(k,t) &= f_k^i + \frac{t}{100}(f_k^{i+1} - f_k^i), \quad k = 1,2,...,L \\
a(k,t) &= a_k^i + \frac{t}{100}(a_k^{i+1} - a_k^i), \quad k = 1,2,...,L
\end{aligned}
\tag{10}
$$

以求取時刻 t 時各諧波的頻率與振幅，其中 100 表示相鄰音框之間的樣本點數，L 是 L^i 和 L^{i+1}的較大者，因此當 L^i 小於 L^{i+1}時，就要把 a_k^i, k= L^i +1,…, L^{i+1}設爲零值。然後，以如下公式計算 $h(t)$，

$$
\begin{aligned}
h(t) &= \sum_{k=1}^{L} a(k,t)\cdot\cos\big(\phi(k,t)\big), \quad 0 \le t < 100 \\
\phi(k,t) &= \phi(k,t-1) + 2\pi\cdot f(k,t)/22,050
\end{aligned}
\tag{11}
$$

其中 $\phi(k,t)$ 表示第 k 個諧波累積到時刻 t 時的相位量，關於初始值 $\phi(k, -1)$，我們令其等於前一音框裡的 $\phi(k,99)$ 以便維持相位的連續性，而當音框編號 i 爲 0 時就以亂數來設定。

關於雜音(noise)信號的合成，我們採取 HNM 文獻上提到的一個作法[16]，就是把雜音當作是 MVF 之後頻率間隔固定爲 100Hz、但振幅會隨時間改變之一些弦波的加總。先依 MVF 決定頻率 index 之下限 KL=MVF / 100，而其上限明顯地是 KU=11,025 / 100，如此，對於第 i 和第 i+1 音框之間時刻 t 的雜音信號樣本 $g(t)$，我們以如下公式來計算，

$$
\begin{aligned}
\mathrm{g}(t) &= \sum_{k=KL}^{KU} b(k,t)\cdot\cos\big(\psi(k,t)\big), \quad 0 \le t < 100 \\
\psi(k,t) &= \psi(k,t-1) + 2\pi\cdot k\cdot 100/22,050
\end{aligned}
\tag{12}
$$

其中 $b(k,t)$ 表示時刻 t 時第 k 個弦波的振幅，其值也是以類似公式(10)之線性內差來求得，$\psi(k,t)$ 表示第 k 個弦波累積到時刻 t 時的相位量，其初始值也是以亂數來設定。最後，將 $h(t)$與 $g(t)$相加，即可得到時刻 t 的合成信號樣本。

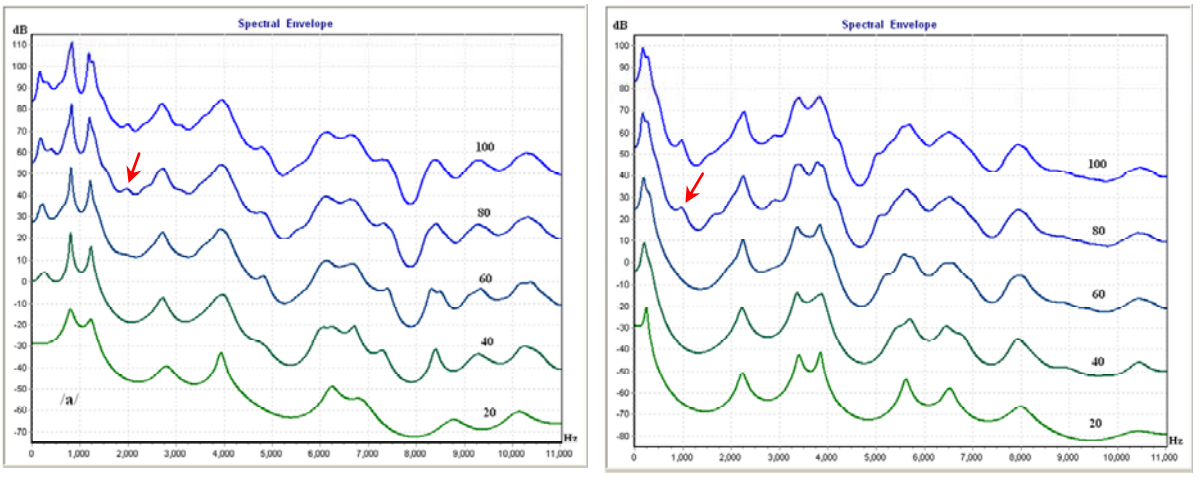
5. 測試實驗

5.1 全極模型之階數

前面 3.2 節提到，以內插得到的反射係數去建造對應的全極模型，再以全極模型的頻率響應作爲導出的頻譜包絡。那麼全極模型的階數應設多大？在此以實驗的方式來觀察不同階數對頻譜包絡曲線的影響，我們先對一個/a/音音框及一個/i/音音框去計算它們的 DCep 之頻譜包絡[8]，然後依序執行 3.1 節和 3.2 節的轉換處理，並且分別建立階數 20、40、60、80、100 之全極模型，再依公式(8)去算出它們的頻譜包絡曲線，畫成圖形後/a/音框和/i/音框各自的 5 種階數的頻譜包絡曲線，分別如圖 6(a)與 6(b)所示。當比較原始的

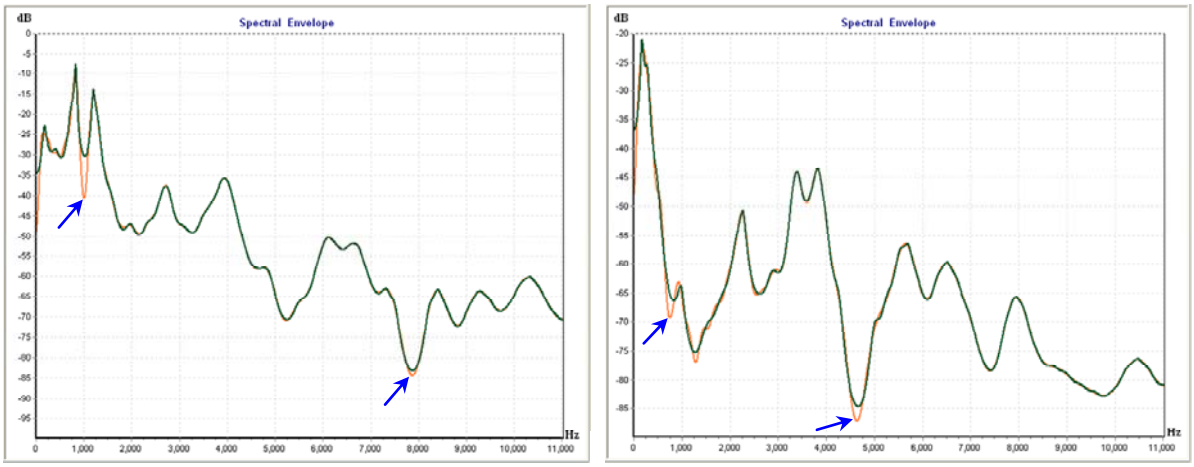
頻譜包絡和 5 種階數的頻譜包絡，我們發現階數要到 80 時，/a/和/i/的兩個音框的頻譜包絡才會足夠細緻，也就是圖 6(a)與 6(b)裡箭頭所指的頻譜隆起才會顯現出來。因此在後面的實驗中，我們就使用階數爲 80 之全極模型。

對於階數 80 之全極模型，我們再進一步觀察它的頻譜包絡和原始的 DCep 頻譜包絡之間的逼近程度，使用相同的/a/和/i/兩個音框，它們的全極模型頻譜包絡的逼近情形分別如圖 7(a)與 7(b)所示，由圖 7(a)與 7(b)可發現，在頻譜包絡曲線的波峰與平緩變化的地方，全極模型的頻譜包絡曲線(深色)幾乎是和 DCep 頻譜包絡曲線(淺色)重疊在一起，只有在頻譜包絡曲線的尖銳波谷的地方，才會顯現出明顯的逼近誤差，如圖 7(a)與 7(b)裡箭頭所指的地方。所以由反射係數所導出的全極模型之頻譜包絡曲線，一般來說可以相當準確地逼近原始的頻譜包絡曲線。



(a) /a/音框的 5 條頻譜包絡　　(b) /i/音框的 5 條頻譜包絡

圖 6　全極模型階數對頻譜包絡曲線的影響



(a) /a/音框的頻譜包絡　　(b) /i/音框的頻譜包絡

圖 7　階數 80 之全極模型頻譜包絡的逼近情形

5.2 共振峰軌跡之走勢

在此我們將一個/a/音音框和一個/i/音音框的反射係數，以逐漸改變比例值的方式(就如公式(9)中間的式子)，內插出一序列 9 個音框的反射係數，然後使用 3.2 節的轉換處理，把反射係數轉換成頻譜包絡，再將 11 個音框(內插 9 個及原始 2 個)的頻譜包絡依序由下往上畫出，結果得到如圖 8 所顯示的圖形。由圖 8 可看出，/a/音音框的 F1(第一共振峰)頻率值會逐漸地往/i/音音框的 F1 頻率值移動，而/a/音框的 F2(第二共振峰)頻率值也會逐漸地往/i/音框的 F2 頻率值移動，就如圖 8 裡的兩條縱向虛線所顯示的，所以經由反射係數作內插，的確可以達成第 2 節裡所提到的頻譜包絡內插之效果。

另外，我們也使用相同的/a/音與/i/音音框，帶入公式(9)去內插出 120 個音框，再設定各音框的基頻值都爲 130Hz，然後使用第 4 節的語音信號再合成方法，去產生出語音信號並且存成音檔，此音檔可從網頁http://guhy.csie.ntust.edu.tw/envlp/ 去下載試聽，我們聆聽之後覺得，該音檔聽起來的確有雙母音/ai/的感受。此外，我們使用 WaveSurfer 軟體去分析此音檔的聲譜，結果得到如圖 9 所示的圖形，從圖 9 我們可以明顯看到前三條共振峰的軌跡，會隨著橫軸(時間軸)逐漸地移動頻率值，而沒有軌跡不連續的情況。

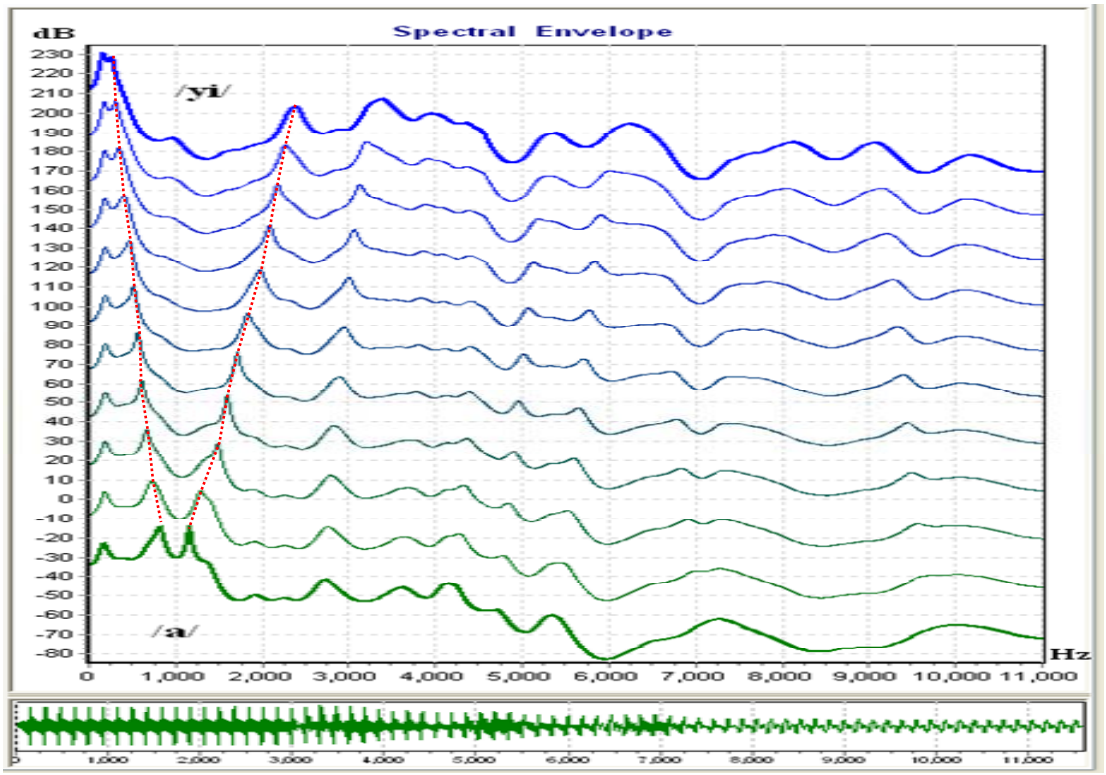


圖 8　經反射係數內插得到之一序列的頻譜包絡

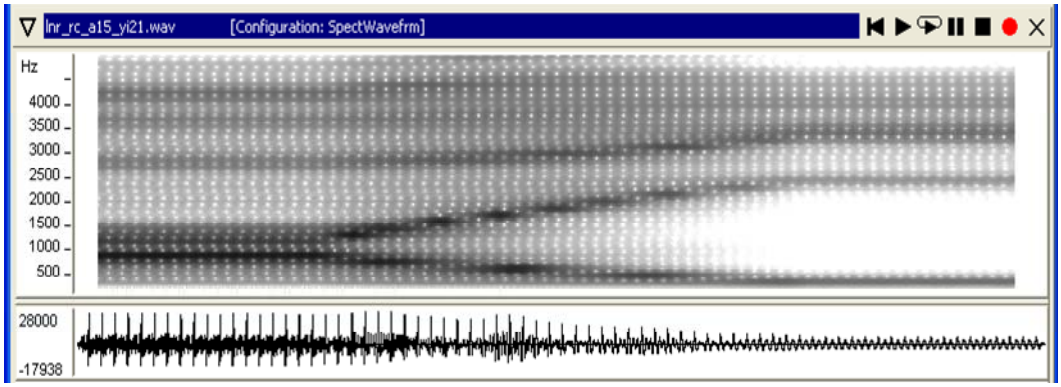


圖 9　經反射係數內插之合成語音的聲譜

爲了作對照，我們也以公式(9)來直接對 DCep 係數作線性內插，內插出的 120 個音框的 DCep 係數，再拿去計算出對應的頻譜包絡[8]，然後使用第 4 節的語音信號再合成方法，去產生出語音信號並且存成音檔。當聆聽這個音檔，會感覺聽到兩個獨立的音素，而無雙母音/ai/的感受。此外，以 WaveSurfer 軟體去分析這個音檔的聲譜時，結果得到如圖 3 所示的圖形，即顯現出淡進淡出之共振峰軌跡的走勢，所以直接對 DCep 係數作內插，並不能得到所想要的頻譜包絡內插之效果。由圖 3 之淡進淡出現象，我們推測其它的倒頻譜係數(如 LPC cepstrum)，直接作內插時也將會發生淡進淡出的情況。

至於 LPC 係數，我們也曾以公式(9)來直接作線性內插，內插出的 120 個音框的 LPC 係數，再以公式(8)計算出對應的頻譜包絡，然後使用第 4 節的語音信號再合成方法，去產生出語音信號。LPC 係數內插出的頻譜包絡，似乎可以讓共振峰軌跡作逐漸的移轉，如圖 10 所示之聲譜，不過 F1 和 F2 的軌跡並不是隨著橫軸作線性的改變，並且聲譜上會出現一些能量突然變高的特黑的點，如圖 10 裡箭頭所指的地方。其原因是 LPC 係數內插出的頻譜包絡，在某些時間點上會發生不穩定的情況，即全極模型有極點跑到 z 平面的單位圓上[15]，而造成很尖銳的頻譜包絡波峰的出現。就因爲這樣，合成出的語音信號，聽起來才會感覺到含有喀嗒聲(click)。

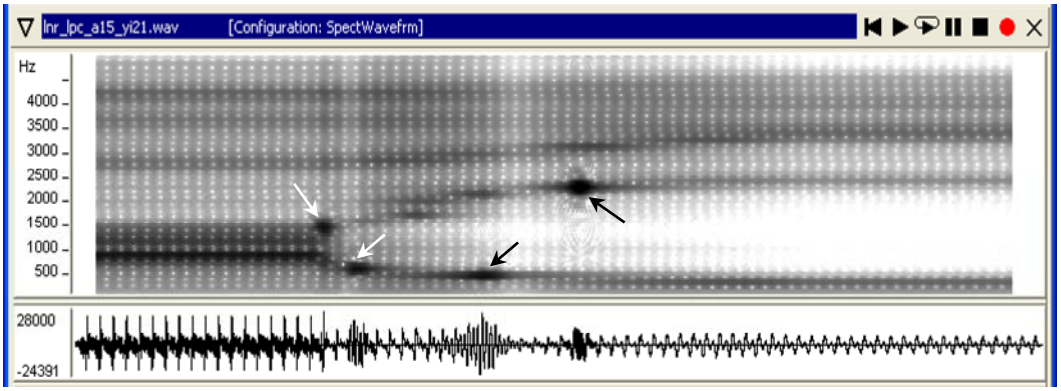


圖 10　經 LPC 係數內插之合成語音的聲譜

6. 結語

關於頻譜包絡的內插，若要達成共振峰軌跡的平滑移動，並不是各種的頻譜參數都能做到。在本篇論文裡，我們提出了一種以反射係數來控制共振峰軌跡作平滑移動的方法，並且說明了反射係數與頻譜包絡之間相互作轉換的實施程序。爲了驗證所提出之方法，我們撰寫了程式作反射係數內插的測試，實驗結果顯示，當對反射係數作線性內插之處理，則 F1 與 F2 的共振峰軌跡會以線性方式來移動。因此，如果需要獲得非線性的共振峰軌跡移動，如 $0.5 + 0.5\cdot\cos(\pi\cdot t/T)$，則可推廣成以非線性的方式來對反射係數作內插，如此就能達成所希望的效果。不過美中不足的是，反射係數所轉換出的頻譜包絡，在包絡曲線的尖銳波谷地方，逼近誤差會顯得較大，這是未來需要再作改進的。

參考文獻

[1] D. T. Chappell and J. H. L. Hansen, “A comparison of spectral smoothing methods for segment concatenation based speech synthesis,” *Speech Communication*, vol. 36, pp. 343–374, 2002.

[2] D. Schwarz and X. Rodet, “Spectral envelope estimation and representation for sound analysis-synthesis,” *Int. Computer Music Conference*, Beijing, China, pp. 351-354, Oct. 1999.

[3] A. Robel and X. Rodet, “Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation,” *Int. Conference on Digital Audio Effects*, Madrid, Spain, pp. 1-6, 2005.

[4] J. D. Markel and A. H. Gray Jr., *Linear Prediction of Speech*, New York: Springer-Verlag, 1976.

[5] S. Imai and Y. Abe, “Spectral envelope extraction by improved cepstral method,” *Electron. and Commun. in Japan*, vol. 62-A, no. 4, pp. 10–17, 1979. (in Japanese)

[6] H. Kawahara, I. Masuda-katsuse, and A. De Cheveign, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds”, *Speech Communication*, vol. 27, pp. 187-207, 1999.

[7] O. Cappé and E. Moulines, “Regularization techniques for discrete cepstrum estimation,” *IEEE Signal Processing Letters*, vol. 3, no. 4, pp. 100–102, 1996.

[8] 古鴻炎、蔡松峰，「基於離散倒頻譜之頻譜包絡估計架構及其於語音轉換之應用」，第二十一屆自然語言與語音處理研討會(ROCLING 2009)，台中，第 151-164 頁，2009。

[9] D. O`Shaughnessy, *Speech Communication: Human and Machine*, 2nd ed., IEEE Press, 2000.

[10] H. R. Pfitzinger, “DFW-based spectral smoothing for concatenative speech synthesis,” *Int.*

Conf. Spoken Language Processing, Jeju, Korea, pp. 1397-1400, 2004.

[11] T. Ezzat, E. Meyers, J. Glass, and T. Poggio, "Morphing spectral envelopes using audio flow," *INTERSPEECH 2005*, Lisbon, Portugal, pp. 2545-2548, 2005.

[12] K. Tokuda, H. Zen, and A. W. Black, “An HMM-based speech synthesis system applied to English,” *IEEE Workshop on Speech Synthesis*, Santa Monica, CA, pp. 227-230, 2002.

[13] 王小川，語音訊號處理(修訂二版)，全華圖書公司，台北，2009。

[14] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, 1993.

[15] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed., Prentice-Hall, 1999.

[16] Y. Stylianou, *Harmonic plus Noise Models for Speech, Combined with Statistical Methods, for Speech and Speaker Modification*, Ph.D. thesis, Ecole Nationale Supèrieure des Télécommunications, Paris, France, 1996.