

本期要目

壹. 學術活動預告—Call for Papers -ACM TALIP

第二頁

貳. 專文—現階段強健性語音辨識技術研究之簡介(林士翔、陳柏琳)

第三~十二頁

第十屆理監事當選名單

理事長：

陳信希(台灣大學資訊工程所教授)

副理事長：

吳宗憲(成功大學資訊工程所教授)

理事：

簡立峰(Google 台灣工程研究所所長)

許聞廉(中央研究院資訊所研究員)

簡仁宗(成功大學資訊工程系教授)

陳光華(台灣大學圖書資訊系副教授)

余明興(中興大學資訊科學系副教授)

曾元顯(台灣師範大學數位媒體中心研究員)

張智星(清華大學資訊工程系副教授)

郭志忠(工研院資通所前瞻技術中心副主任)

曾淑娟(中央研究院語言所副研究員)

張景新(暨南國際大學資訊工程系助理教授)

高照明(台灣大學外文系助理教授)

陳柏琳(台灣師範大學資訊工程系副教授)

盧文祥(成功大學資訊工程系助理教授)

候補理事：

廖元甫(台北科技大學電子通訊所助理教授)

王逸如(交通大學電信工程系副教授)

沈家麟(台達電子工業(股)公司研發部經理)

監事：

陳信宏(交通大學電信工程系教授)

陳克健(中央研究院資訊所研究員)

王小川(清華大學電機工程系教授)

張俊盛(清華大學資訊工程系教授)

黃居仁(中央研究院語言所研究員)

Web 2.0 技術與應用研討會

2007 年資訊檢索研討會謹訂於 96/12/19 (三)假中研院資訊所舉行，本次會議的主題為「Web 2.0 技術與應用」，會中邀請到中研院研究員許聞廉教授、清華大學張俊盛教授、中正大學吳昇教授及資策會電子商務所副所長楊仁達博士等來賓發表演說，演講主題及報名方式將於近日公布於學會網頁，屆時歡迎有興趣者踴躍報名參加。

第七屆博碩士論文獎得獎名單

博士論文獎：

優等獎一名：獲獎金二萬元及獎狀一紙

發表者：鍾曉芳(台灣大學語言學研究所)

論文題目：以語料庫驅動之隱喻源域界定研究

指導教授：安可思 教授

碩士論文獎：

優等獎：從缺

佳作獎三名：各獲獎金伍仟元及獎狀一紙

發表者：梁玉婷(台灣大學資訊工程學所)

論文題目：意見問答系統中問題分析與答案區段選取之研究

指導教授：陳信希 教授

發表者：梁仲豪(成功大學資訊工程系)

論文題目：個人習語之萃取與生成於說話風格模型化之研究

指導教授：吳宗憲 教授

發表者：林士翔(台灣師範大學資訊教育所)

論文題目：數據擬合與分群方法於強健語音特徵提取之研究

指導教授：葉耀明 教授、陳柏琳 教授

ACM Transactions on Asian Language Information Processing
Special Issue on
Recent Advances in Asian Language Spoken Document Retrieval
(<http://chinese.csie.ncku.edu.tw/CallForPaper/>)

With the rapid expansion of audio media sources such as radio, TV, lectures, and telephony recordings, there is an increasing demand for automatic indexing and retrieval of spoken documents. Spoken Document Retrieval (SDR) is essentially the task of retrieving excerpts from a large collection of spoken documents based on a user's request. Over the past several years, SDR has emerged as an interdisciplinary field combining automatic speech recognition, natural language processing, and information retrieval. SDR has benefited from advances in speech and language processing as well as from the availability of large spoken databases. With its focus on fundamental research in natural spoken language processing, SDR is key to applications such as voice search, voice surveillance, voice data mining and call center automation.

We invite papers on recent advances in Asian language SDR research. Submissions must not have been previously published, with the exception that substantial extensions of conference papers will be considered. Of particular interest are studies that directly address problems in Asian spoken languages, but new and general methods in SDR tested on Asian languages are also welcome. Topics of interest include:

- Speech recognition and adaptation algorithms
- Natural language processing for SDR
- Topic detection and tracking, routing
- Clustering and classification
- Indexing and retrieval
- Summarization and information extraction
- Understanding, organization and visualization
- Cross-language and multilingual SDR
- Query-by-example SDR
- Spoken dialogue systems for accessing spoken archives
- Spoken document corpora
- Performance, compression, scalability, architectures, efficiency
- Voice data mining and machine learning for SDR
- SDR in creative applications

SUBMISSION PROCEDURE:

Papers should be formatted following the style guidelines for the ACM Transactions on Asian Language Information Processing

(http://www.acm.org/pubs/talip/authors.html#STYLE_GUIDELINES).

Please submit papers in PDF format using the Web-based Submission System Manuscript Central (<http://mc.manuscriptcentral.com/talip>). In order to identify the submission is for the special issue, please indicate "This paper is submitted to the Special Issue on Asian Language SDR" under the paper title on the first page of the paper.

SCHEDULE:

Submission deadline	1 January 2008
Completion of first round of reviews	15 April 2008
Revisions Due	15 May 2008
Notification of final acceptance	15 August 2008
Final manuscript due	15 September 2008
Publication date	December 2008

GUEST EDITORS:

Prof. Chung-Hsien Wu (National Cheng Kung University, Taiwan)
Dr. Haizhou Li (Institute for Infocomm Research, Singapore)

chwu@csie.ncku.edu.tw
hli@i2r.a-star.edu.sg

現階段強健性語音辨識技術研究之簡介

林士翔 陳柏琳

國立台灣師範大學資訊工程學系

{ shlin, berlin }@csie.ntnu.edu.tw

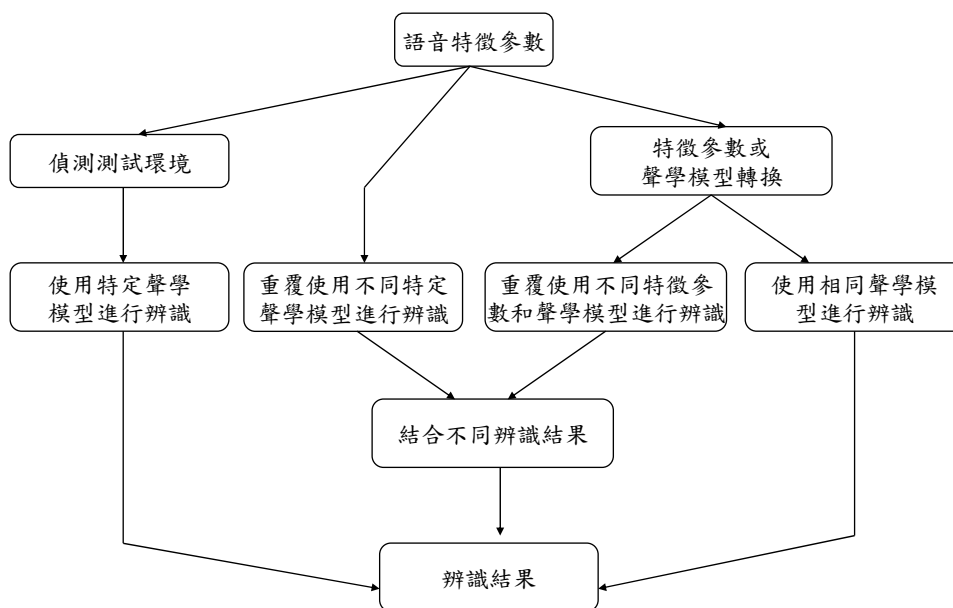
一、前言

語音長久以來一直是人類最自然且最容易使用的溝通媒介，無庸置疑地，語音勢必會扮演著未來人類與各種智慧型電子設備間最主要的人機互動媒介，因此自動語音辨識(Automatic Speech Recognition, ASR)技術將會是扮演關鍵且重要的角色[1]。此外，隨著資訊科技的發展，日常生活中充斥著各式各樣的多媒體影音資訊，例如廣播電視節目、語音信件、演講錄影和數位典藏等，然而大部份的多媒體資訊中，語音可以說是最具語意的主要內涵之一，因此，語音辨識技術對多媒體資訊處理也扮演著相當重要的角色[2]。自動語音辨識的挑戰在於給定一段語音訊號，如何讓電腦能夠快速且正確的辨識出語音訊號的內容，並將其轉換成一連串詞序列(Word Sequence)。

現今語音辨識系統的辨識效能，根據語音辨識任務的複雜度不同，辨識效能亦有所差距，例如語者相關(Speaker-Dependent)的小詞彙獨立字詞辨識(Small Vocabulary Isolated Speech Recognition, SVISR)的辨識系統，通常能有 95%以上的辨識正確率；若辨識任務是語者獨立(Speaker-Independent)的大詞彙連續語音辨識(Large Vocabulary Continuous Speech Recognition, LVCSR)，則辨識正確率約莫達 80%左右。但是如果測試環境是在非常吵雜的情境下進行，系統辨識效能則可能會大幅度降，主要是因為環境中複雜因素的影響，造成訓練環境與測試環境存在環境不匹配(Environmental Mismatch)的差異，其中干擾的因素可從許多不同層面探討[3]，例如語者發音結構差異、語音訊號輸入的通道效應、加成性噪音(Additive Noise)、頻譜傾斜(Spectral Tilt)或其他語者干擾等。正因如此，語音強健(Robust Speech)技術長久以來一直被視為重要的研究課題[4-6]，其目的主要希望藉由對語音訊號本身、語音特徵參數或是聲學模型參數做適當的處理，以減緩雜訊干擾的影響情形、降低訓練環境與測試環境不匹配的情形，或是加強語音訊號或語音特徵參數本身的強健性，進而提高辨識系統的辨識效能。以下，本文將針對當前語音強健技術的相關技術發展作簡介。

二、語音強健技術

爲了解決環境不匹配問題，目前在語音強健技術研究上大致可分二個方向[7]：第一個方向是針對不同測試環境，分別訓練出專用的聲學模型(Acoustic Models)，另外一個研究方向是針對語音特徵參數或聲學模型做適當的處理，例如特徵參數轉換(Transformation)、增益(Enhancement)或聲學模型調適(Adaptation)等方法，以降低雜訊造成訓練環境與測試環境不匹配的影響。此二種研究方向示意圖如圖一所示。



圖一、 解決訓練環境與測試環境不匹配的辨識方法

2.1 偵測測試環境，使用特定聲學模型進行辨識

主要是訓練各種不同測試環境情況下的聲學模型，例如特定語者(Speaker-dependent)、特定性別(Gender-dependent)、特定噪音(Noise-dependent)、或特定通道(Channel-dependent)等，然後再利用與測試環境相符的聲學模型進行辨識。但此方法必須有個前提假設，在進行辨識時，必須能夠線上即時決定辨識環境是什麼環境，進而決定要使用哪一組特定聲學模型進行辨識。若沒有辦法事先決定測試環境爲何時，可分別使用各個聲學模型進行一次辨識，再將各種聲學模型所辨識出的辨識結果結合，找出最佳的辨識結果[8]。此研究方向的優點是當測試環境可以正確地被識別出(Identified)時，使用與測試環境相匹配的聲學模型進行辨識，那麼應可得到良好的辨識效能。但其缺點是因爲現實環境中影響語音辨識效能的因素太多，並無法在訓練聲學模型時考慮到所有的可能情況；因此在進行語音辨識時，若無法找到相對應環境的聲學模型，那麼辨識效能還是無法提昇。

2.2 語音特徵參數或聲學模型轉換

在此類研究方向中，不論是在聲學模型訓練階段或語音辨識階段都只利用一組通用(Universal)聲學模型，主要有三類作法：

(I) 語音強化技術(Speech Enhancement)

語音強化技術目的在於提升語音訊號本身的品質，通常是假設語音訊號與雜訊訊號二者在統計上是不相關(Unrelated)，希望能由觀察到的雜訊語音(Noisy Speech)重建還原出原本的乾淨語音(Clean Speech)訊號，常見的技術有頻譜消去法(Spectral Subtraction, SS)[9]、維爾濾波器(Wiener Filter, WF)[6]或卡爾曼濾波器(Kalman Filter)[10]等。

(II) 強健性語音特徵(Robust Speech Features)

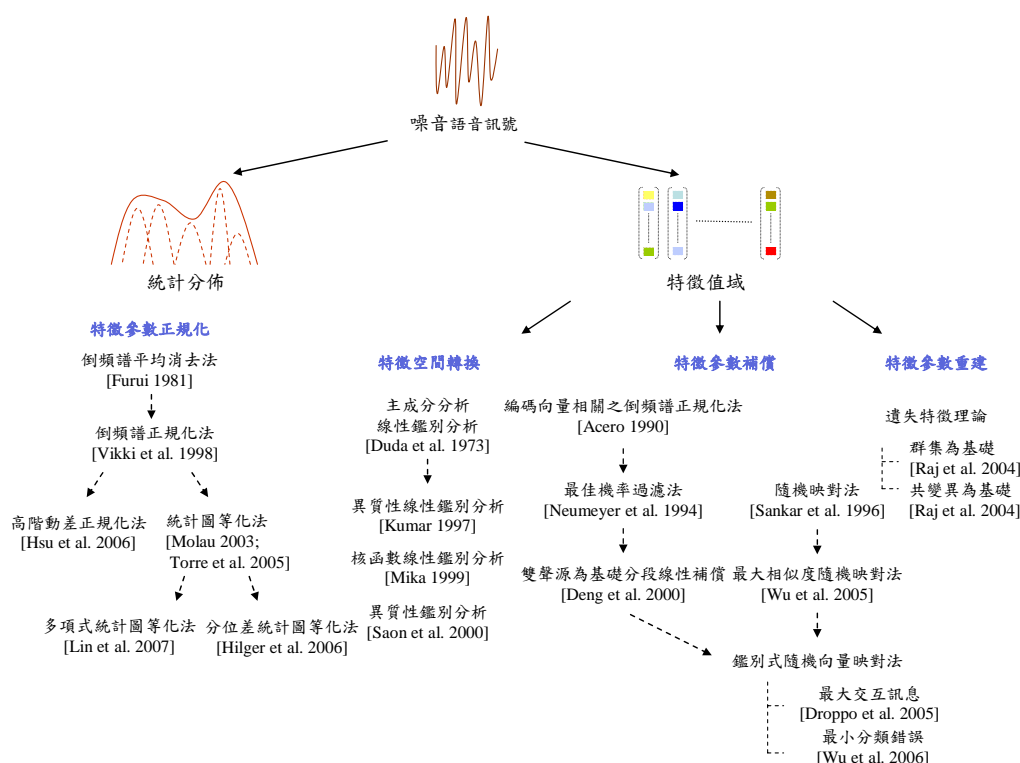
主要是從語音訊號中擷取出較不易受到環境變化干擾而失真的強健性語音特徵參數或是對語音特徵參數進行補償，目前的研究議題又可再進一步分成數個研究方向，包括語音的特徵參數正規化法 (Feature Normalization)、特徵參數轉換法 (Feature Transformation)、特徵參數補償法 (Feature Compensation) 與特徵參數重建法 (Feature Reconstruction) 等，此類作法為本文介紹重點，近年來較被廣泛討論的方法將於下一章節描述。

(III) 聲學模型調適(Model Adaptation)

藉由少量在測試環境中收錄的調適語料(Adaptation Data)來調整通用聲學模型中的機率分布參數，期望調適後的模型可以適用於新的環境，以降低環境不匹配的現象。常見的技術有最大事後機率法則(Maximum a Posteriori, MAP)[11-12]、最大相似度線性回歸法(Maximum Likelihood Linear Regression, MLLR)[13-14]及平行模型合併法(Parallel Model Combination, PMC)[15-17]等。

縱觀上述之三種研究方向，又以聲學模型調適可獲得較佳的辨識效能，因為藉由調適語料與對應的正確轉譯文字(Correct Transcripts)的使用，直接調整聲學模型參數，以降低由雜訊所產生環境不匹配的不確定性(Uncertainty)統計特性。但是此作法相較於其他作法，必需有額外在測試環境下所收錄的語音與正確轉譯文字的調適資料，且需花費額外相當的運算時間以進行聲學模型參數調整。另一方面就語音強化技術而言，其主要目的是消除雜訊並改善語音的品質，例如提昇語音的訊噪比(Signal-to-Noise Ratio, SNR)，但此舉通常並不一定保證能可以提高語音辨識效能。相較於前述作法，強健性語音特徵不僅可有效地提昇辨識效能，而且通常只需額外短暫的運算時間即可完成。

三、強健性語音特徵



圖二、強健性語音技術分類圖

目前強健性語音特徵的主要研究方向，大致上可從二個方向進行：第一類是直接從雜訊語音特徵參數的特徵值域(Feature Domain)進行特徵參數補償或特徵參數轉換或重建，此類研究方向通常假設有關於雜訊干擾語音訊號的先備知識(Prior Knowledge)或是假設乾淨語音特徵參數與雜訊語音特徵參數間存在著某種固定的關係，所以在訓練階段可以事先求算此一關係，而在測試階段利用此一關係進行補償；另一種研究方向是利用一些較不容易受雜訊干擾而有所影響的語音特徵參數的統計分布特性(Distribution Characteristics) 當作是特徵參數補償或是特徵參數轉換的依據，此研究方向相較於前者只需額外的短暫運算時間即可獲得良好的辨識效能，分類圖如圖二所示，下列小節將分述近年來一些較廣為討論的方法。

3.1 特徵參數正規化法

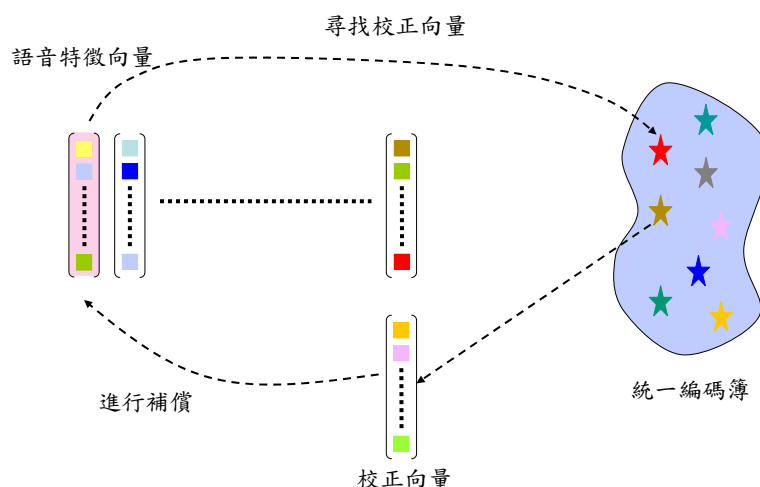
語音特徵參數正規化法，期望藉由對語音特徵參數的正規化過程中，進一步移除雜訊干擾的影響，目前最常見的方法之有倒頻譜平均消去法(Cepstral Mean Substraction, CMS)[18]與倒頻譜正規化法(Cepstral Mean and Variance Normalization, CMVN)[19]，其分別針對語音特徵參數的第一和第二階動差(Moment)進行正規化，藉由正規化的過程，除了移除通道效應所造成的影響，同時倒頻譜正規化法更可進一步降低了每個維度間語音特徵機率分布的差異程度，以減少環境不匹配對語音特徵參數所造成的不良影響。但是由於倒頻譜平均消去法與倒頻譜正規化

法本身方法線性關係的限制，造成只能補償因受雜訊干擾影響所產生的線性失真部份，對於非線性失真部份的補償效果有限，因此便有許多學者嘗試提出許多不同的補償方法，試圖解決因雜訊干擾影響對語音特徵參數所產生的失真情形，例如針對語音特徵參數的第三階動差進行正規化[20]或對語音特徵參數更高階動差進行正規化[21-22]。此外，亦有學者嘗試將已經在影像處理中行之有年的統計圖等化法(Histogram Equalization, HEQ)應用於語音辨識之特徵參數正規化[7, 23-26]，統計圖等化法除了試圖去匹配訓練語料與測試語料之語音特徵參數的平均數和變異數之外，更企圖讓訓練語料和測試語料的統計分布特性能夠相同(Identical)，換言之，統計圖等化法是針對語音特徵參數的每一階動差進行正規化，其作法是藉由將測試語料的累積密度函數(Cumulative Density Function, CDF)值應對至由訓練語料所統計出來的參考分布的累積密度函數值，藉由此匹配轉換過程，降低測試語料與訓練語料由於環境因素影響所造成統計特性不匹配的現象，實驗結果證實統計圖等化法對於提升辨識效果有很明顯的幫助[23-25]。有別於傳統查表式統計圖等化法或分位差統計圖等化法需耗費的大量記憶體資源與處理器運算時間的缺點，新近亦有研究提出以多項式迴歸模型，透過少量的多項式係數與多項式函數的運用，迅速的將測試語句語音特徵參數每一維的統計分布轉換至先前已從訓練語句中定義好的參考分布[26]。

3.2 特徵空間轉換法

因為語音特徵參數容易受到雜訊干擾而產生變化，因此便有研究學者嘗試不同的轉換方法，期望找出更具強健性的語音特徵，且不易受到雜訊或者通道效應影響，其中以資料相關線性語音特徵空間轉換(Data-Driven Linear Feature Transform)最廣為探討，資料相關線性語音特徵空間轉換的好處除了藉由統計訓練語料的統計資訊，自動地找出特徵空間中重要的基底向量，使得經轉換後的語音特徵參數能只保留具較大變異或者有鑑別力的特徵成分，並且能進一步去除多餘(Redundant)的維度，常見的方法有主成分分析(Principal Component Analysis, PCA)[27]、線性鑑別分析(Linear Discriminant Analysis, LDA)[27]、異質性線性鑑別分析(Heteroscedastic Linear Discriminant Analysis, HLDA)[28-29]、異質性鑑別分析(Heteroscedastic Discriminant Analysis, HDA)[30]。其中線性鑑別分析是假設所有類別特徵向量的分布變異是相同的；而異質性線性鑑別分析與異質性鑑別分析則是打破這樣的假設，允許類別間的分布變異可以不同。同時，也有研究嘗試以核函數線性鑑別分析(Kernel Linear Discriminant Analysis, Kernel LDA) [31]對語音特徵參數做進一步處理，利用核函數(Kernel Function)將語音特徵參數投影(Project)到高維度的特徵空間，再作線性鑑別分析，以解決在原本特徵空間中可能存在著非線性鑑別的問題。

3.3 特徵參數補償法



圖三、編碼為基礎之倒頻譜向量補償法示意圖

在語音特徵參數補償法中，近年來被廣為研究與討論的議題為以編碼為基礎之倒頻譜向量補償法(Codebook-based Cepstral Compensation)，在此類方法中，主要的精神是利用向量量化編碼技術對語音特徵參數建立一組統一編碼簿(Universal Codebook)，且統一編碼簿裡的每一個向量量化編碼區域(Vector Quantization Region)都有一組對應的補償向量，用來當作是將雜訊語音特徵參數還原至乾淨語音特徵參數的補償參考值。在測試時，對於每個音框而言，都需至統一編碼簿中找出一個最相似的向量量化編碼區域，再利用該編碼區域所對應的補償向量進行補償，整體示意圖如圖三所示。目前常見的方法有編碼詞相關之倒頻譜正規化法(Codeword Dependent Cepstral Normalization, CDCN)[3]、機率最佳化過濾法(Probabilistic Optimum Filtering, POF)[32]與雙聲源為基礎分段線性補償(Stereo-based Piecewise Linear Compensation, SPLICE)[33-35]或隨機特徵向量對映法(Stochastic Vector Mapping, SVM)[36-39]等。

3.4 特徵參數重建法

語音特徵參數重建法又以遺失特徵理論(Missing Feature Theory, MST)為代表，遺失特徵理論可作用在前端語音特徵擷取[40-42]或是後端語音解碼[43]上，不論是做在前端特徵擷取或後端語言解碼，基於遺失特徵理論的語音特徵參數重建法基本上都需包含二個步驟：第一步是決定哪些語音特徵參數是可靠(Reliable)，哪些是不可靠(Unreliable)或遺失(Missing)的，常見的作法包括估測每個音框中頻譜的訊噪比[44]、利用語音訊號能量強度當為準則[45]、利用貝氏分類器(Bayesian Classifier)進行判別[40]或是利用一些聽覺感知(Perceptual)特性或音韻(Prosodic)資訊進行判別[46]；第二步則是針對不可靠的語音特徵參數進行參數重建。若作用在前端語音特徵擷取上，其又可視為特徵向量設算法(Feature Vector Imputation)，最主要的目

的是利用可靠的語音特徵參數(區域)重建不可靠語音特徵參數(區域)的特徵值，重建的方法可利用事前先用乾淨語料所先訓練好的高斯混合模型進行不可靠特徵參數的估測[40]；然而若遺失特徵理論作用在後端語音解碼上，除了可用特徵向量設算法估測不可靠語音特徵參數的特徵值外，亦可使用邊際化法(Marginalization)，並不用重建任何語音特徵參數，而是直接利用可靠的語音特徵參數進行相似度估測[40]。

四、結論

強健性語音辨識技術在語音辨識研究領域裡，一直是一個非常重要的研究議題，唯有妥善解決雜訊干擾語音所造成的失真問題，語音辨識技術方能慢慢地被大眾所接受。目前的研究方向大致可分為從語音特徵參數的特徵值域或是利用其所對應的統計分布特性進行處理。然而，就第一種研究方向卻存在一個潛在的隱藏問題，因為雜訊干擾語音訊號的影響並非絕對是一對一的線性關係，所以可能因某些非預期的因素影響，造成特徵參數補償或特徵參數轉換的效果不佳，或更進一步使得辨識效能驟降；此外，以語音特徵參數的統計特性出發相較於第一種研究方向，只需額外的短暫運算時間即可獲得良好的補償效果，但有些方法往往會事先對語音特徵參數的統計分布做一些限制或假設，例如假設語音特徵參數的分布是高斯分布等，然而此類假設並非完全正確，因此可能使得方法的補償效果有所侷限，此外，雜訊干擾語音訊號除了會產生統計特性不匹配的問題外，因其本身的隨機特性，同時也會對語音訊號加入了不確定性(Uncertainties)，而此研究方向只能有效處理統計特性不匹配的問題，卻無法解決由雜訊干擾所產生的不確定性問題，因此若能結合二種研究方向的優點，勢必能更有效的提升語音辨識效能[47]。

五、參考文獻

- [1] Juang, B. H. and S. Frui (2000), "Automatic Recognition and Understanding of Spoken Language — A First Step Toward Natural Human-Machine Communication," *Proceedings of the IEEE*, vol. 88(8): pp. 1142-1165.
- [2] Lee, L. S. and B. Chen (2005), "Spoken Document Understanding and Organization," *IEEE Signal Processing Magazine (IEEE SPM)*, vol. 22(5): pp. 42-60.
- [3] Acero, A. (1990), "Acoustical and Environmental Robustness for Automatic Speech Recognition," *Ph. D. Dissertation*, Carnegie Mellon University.
- [4] Gong, Y. (1995), "Speech Recognition in Noisy Environments: A Survey," *Speech Communication*, vol. 16(3): pp. 261-291.
- [5] Junqua, J. C. and J. P. Haton (1996), "Robustness in Automatic Speech Recognition," Norwell, MA:Kluwer.
- [6] Huang, X., A. Acero and H. W. Hon (2001), "Spoken Language Processing: A Guide to Theory, Algorithm and System Development," Upper Saddle River, NJ, USA, Prentice Hall PTR.

- [7] Molau, S. (2003), "Normalization in the Acoustic Feature Space for Improved Speech Recognition," *Ph. D. Dissertation*, RWTH Aachen University.
- [8] Fiscus, J. (1997), "A Post-Processing System to Yield Reduced Word Error Rates: Recognizer Output Voting Error Reduction (ROVER)," *IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU'97)*, Santa Barbara, California.
- [9] Boll, S. F. (1979), "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27(2): pp. 113-120.
- [10] Koo, B., J. Gibson and A. Gray. (1989), "Filtering of Colored Noise for Speech Enhancement and Coding," *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP '89)*, Glasgow, Scotland.
- [11] Gauvain, J. L. and C. H. Lee (1994), "Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," *IEEE Transaction on Speech and Audio Processing*, vol. 2(2): pp. 291-297.
- [12] Huo, Q., C. Chany and C. H. Lee (1995), "Bayesian Adaptive Learning of the Parameters of Hidden Markov Model for Speech Recognition," *IEEE Transaction on Speech and Audio Processing*, vol. 3(4): pp. 334-345.
- [13] Leggetter, C. J. and P. C. Woodland (1995), "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models," *Computer Speech and Language*, vol. 9: pp. 171-185.
- [14] Gales, M. J. F. (1998), "Maximum Likelihood Linear Transformations for HMM-based Speech Recognition," *Computer Speech and Language*, vol. 12(2): pp. 75-98.
- [15] Gales, M. J. F. and S. J. Young (1995), "Robust Speech Recognition in Additive and Convolutional Noise Using Parallel Model Combination," *Computer Speech and Language*, vol. 9: pp. 289-307.
- [16] Gales, M. J. F. and S. J. Young (1996), "Robust Continuous Speech Recognition using Parallel Model Combination," *IEEE Transaction on Speech and Audio Processing*, vol. 4(5): pp. 352-359.
- [17] Hung, J. W., J. L. Shen, and L. S. Lee (2002), "New Approaches for Domain Transformation and Parameter Combination for Improved Accuracy in Parallel Model Combination (PMC) Technologies," *IEEE Transactions on Speech and Audio Processing*, vol. 9(8): pp. 842-855.
- [18] Furui, S. (1981), "Cepstral Analysis Techniques for Automatic Speaker Verification," *IEEE Transaction on Acoustic, Speech and Signal Processing*, vol. 29(2): pp. 254-272.
- [19] Vikki, A. and K. Laurila (1998), "Segmental Feature Vector Normalization for Noise Robust Speech Recognition," *Speech Communication*, vol. 25: pp. 133-147.
- [20] Suk, Y. H., S. H. Choi and H. S. Lee (1999), "Cepstrum Third-Order Normalisation Method for Noisy Speech Recognition," *Electronics Letters*, vol. 35(7): pp. 527-528.
- [21] Hsu, C. W. and L. S. Lee (2004), "Higher Order Cepstral Moment Normalization (HOCMN) for Robust Speech Recognition," *IEEE International Conference on*

- Acoustic, Speech and Signal Processing (ICASSP '04)*, Quebec, Canada.
- [22] Hsu, C. W. and L. S. Lee (2006), "Extension and Further Analysis of Higher Order Cepstral Moment Normalization (HOCMN) for Robust Features in Speech Recognition," *Interspeech'2006 - 9th International Conference on Spoken Language Processing (ICSLP)*, Pittsburgh, Pennsylvania.
- [23] Dharanipragada, S. and M. Padmanabhan (2000), "A Nonlinear Unsupervised Adaptation Technique for Speech Recognition," *Interspeech'2000 - 6th International Conference on Spoken Language Processing (ICSLP)*, Beijing, China.
- [24] Torre, A., A. M. Peinado, J. C. Segura, J. L. Perez-Cordoba, M. C. Benitez, and A. J. Rubio (2005), "Histogram Equalization of Speech Representation for Robust Speech Recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 13(3): pp. 355-366.
- [25] Hilger, F. and H. Ney (2006), "Quantile Based Histogram Equalization for Noise Robust Large Vocabulary Speech Recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14(3): pp. 845-854.
- [26] Lin, S. H., Y. M. Yeh and B. Chen (2007), "A Comparative Study of Histogram Equalization (HEQ) for Robust Speech Recognition," *International Journal of Computational Linguistics and Chinese Language Processing*, vol. 12(2): pp. 217-238.
- [27] Duda, R. O. and P. E. Hart (1973), "Pattern Classification and Scene Analysis," New York, John Wiley and Sons.
- [28] Kumar, N. (1997), "Investigation of Silicon-Auditory Models and Generalization of Linear Discriminant Analysis for Improved Speech Recognition," *Ph. D. Dissertation*, John Hopkins University.
- [29] Gales, M. J. F. (2002), "Maximum Likelihood Multiple Subspace Projections for Hidden Markov Models," *IEEE Transactions on Speech and Audio Processing*, vol. 10(2): pp. 37-47.
- [30] Saon, G., M. Padmanabhan, R. Gopinath and S. Chen (2000), "Maximum Likelihood Discriminant Feature Spaces," *IEEE International Conference on Acoustics, Speech, Signal processing (ICASSP '00)*, Istanbul, Turkey.
- [31] Mika, S. (1999), "Fisher Discriminant Analysis With Kernels," *IEEE International Workshop on Neural Networks for Signal Processing (NNSP 1999)*, Madison, Wisconsin.
- [32] Neumeyer, L. and M. Weintraub (1994), "Probabilistic Optimum Filtering for Robust Speech Recognition," *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP '94)*, Albuquerque, New Mexico.
- [33] Deng, L., A. Acero, M. Plumpe and X. Huang (2000), "Large Vocabulary Speech Recognition under Adverse Acoustic Environments," *Interspeech'2000 - 6th International Conference on Spoken Language Processing (ICSLP)*, Beijing, China.
- [34] Droppo, J., L. Deng and A. Acero. (2001), "Evaluation of the SPLICE Algorithm on the Aurora2 Database," *Interspeech'2001 - 7th European Conference on Speech Communication and Technology (Eurospeech)*, Aalborg, Denmark.

- [35] Droppo, J., L. Deng and A. Acero (2002), "Evaluation of SPLICE on the Aurora 2 and 3 Tasks," *Interspeech'2002 - 7th International Conference on Spoken Language Processing (ICSLP)*, Denver, Colorado.
- [36] Wu, J., Q. Huo and D. Zhu (2005), "An Environment Compensated Maximum Likelihood Training Approach based on Stochastic Vector Mapping," *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP '05)*, Philadelphia, Pennsylvania.
- [37] Sankar A. and C. H. Lee (1996), "A Maximum Likelihood Approach to Stochastic Matching for Robust Speech Recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 4: pp. 190-202.
- [38] Wu, J. and Q. Huo (2006), "An Environment-Compensated Minimum Classification Error Training Approach Based on Stochastic Vector Mapping," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14(6): pp. 2147-2155.
- [39] Droppo, J. and A. Acero (2005), "Maximum Mutual Information SPLICE Transform for Seen and Unseen Conditions," *Interspeech'2005 - 9th European Conference on Speech Communication and Technology (Eurospeech)*, Lisbon, Portugal.
- [40] Raj, B., M. L. Seltzer and R. M. Stern (2004), "Reconstruction of Missing Features for Robust Speech Recognition," *Speech Communication*, vol. 43(4): pp. 275-296.
- [41] Raj, B. (2000), "Reconstruction of Incomplete Spectrograms for Robust Speech Recognition. ECE Department. Pittsburgh," *Ph. D. Dissertation*, Carnegie Mellon University.
- [42] Hamme, H. V. (2004), "Robust Speech Recognition Using Cepstral Domain Missing Data Techniques and Noisy Mask," *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP '04)*, Quebec, Canada.
- [43] Cooke, M., P. D. Green, L. B. Josifovski and A. Vizinho (2001), "Robust Automatic Speech Recognition with Missing and Uncertain Acoustic Data," *Speech Communication*, vol. 34: pp. 267-285.
- [44] Vizinho, A., P. D. Green, M. Cooke and L. B. Josifovski (1999), "Missing Data Theory, Spectral Subtraction and Signal-to-Noise Estimation for Robust ASR," *Interspeech'1999 - 6th European Conference on Speech Communication and Technology (Eurospeech)*, Budapest, Hungary.
- [45] EL-Maliki, M. and A. Drygajlo (1999), "Missing Features Detection and Handling for Robust Speaker Verification," *Interspeech'1999 - 6th European Conference on Speech Communication and Technology (Eurospeech)*, Budapest, Hungary.
- [46] Palomaki, K. J., G. J. Brown and D. L. Wang (2004), "A Binaural Processor for Missing Data Speech Recognition in the Presence of Noise and Small-Room Reverberation," *Speech Communication*, vol. 43(4): pp. 361-378.
- [47] Lin, S. H., Y. M. Yeh and B. Chen (2007), "Investigating the Use of Speech Features and their Corresponding Distribution Characteristics for Robust Speech Recognition," *IEEE workshop on Automatic Speech Recognition and Understanding (ASRU 2007)*, Kyoto, Japan.