

本期要目

- 壹. 網路社群服務計算暨探勘技術研討會 第二~三頁
貳. ACL-IJCNLP 2009 Preliminary Call for Papers 第四頁
參. 專文-如何讓電腦說話? - 論語音合成技術(賴玟杏) 第五~十二頁

2009 International Conference Listings

ACL-IJCNLP 2009

Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th - International Joint Conference on Natural Language Processing

- Conference Date: August 2-7, 2009
- Submission Deadline: February 22, 2009
- Location: Singapore
- <http://www.acl-ijcnlp-2009.org/>

ACM-SIGIR-2009

The 32nd Annual International ACM SIGIR Conference

- Conference Date: July 19-23, 2009
- Submission Deadline: January 19, 2009
- Location : Boston, MA, USA
- <http://sigir2009.org/>

AIRS-2009

The 5th Asia Information Retrieval Symposium

- Conference Date : October 21-23, 2009
- Submission Deadline: April 19, 2009
- Location: Sapporo, Hokkaido, Japan
- <http://www-kb.ist.hokudai.ac.jp/AIRS2009/>

CLSW -2009

The 10th Chinese Lexical Semantics Workshop

- Conference Date: July 27-31, 2009
- Submission Deadline: March 20, 2009
- Location: Yantai, China
- <http://www.chineselex.org/>

ICASSP-2009

The 34th International Conference on Acoustics, Speech, and Signal Processing

- Conference Date: April 19 - 24, 2009
- Submission Deadline: September 29, 2008
- Location: Taipei, Taiwan
- <http://icassp09.com/>

Interspeech-2009

The 10th International Conference on Spoken Language Processing

- Conference Date: September 6-10, 2009
- Submission Deadline: April 17, 2009
- Location: Brighton, UK
- <http://www.interspeech2009.org/>

MT Summit -2009

The 12th Machine Translation Summit

- Conference Date: August 26-30, 2009
- Submission Deadline: April 28, 2009
- Location: Ottawa, Ontario, Canada
- <http://summitxii.amtaweb.org/>

NAACLHLT-2009

North American Chapter of the Association for Computational Linguistics - Human Language Technology Conference

- Conference Date: May 31 - June 5, 2009
- Submission Deadline: December 1, 2008
- Location: Boulder, Colorado, USA
- <http://clear.colorado.edu/NAACLHLT2009/>

網路社群服務計算暨探勘技術研討會

Workshop on Community-based Web Service Computing and Mining

http://www.aclclp.org.tw/doc/IR_08.htm

網路社群又稱為虛擬社群(virtual community)，是一群有共同興趣的使用者在網際網路互動之後，產生的社會群體。網路社群的形式包括了早期的電子佈告欄、論壇、留言板，或是近期才出現的部落格、維基百科。透過社群成員彼此在同一個網路社群進行資訊的交流與分享，各式各樣的網路社群不斷地林立和擴大。近年來，許多網路社群服務研究和重要應用議題已逐漸受到學術界和產業界重視。中央研究院資訊科學研究所、國立台灣大學資訊工程學系、國立成功大學資訊工程學系與中華民國計算語言學學會特別舉辦本次研討會，邀請國內外相關學者專家進行觀念和技術交流，包含網路社群服務計算暨探勘技術、機器學習技術、資訊擷取技術、資訊整合技術等重要研究議題。此研討會係繼 2002 年「資訊自動分類技術研討會」、2003 年「資訊檢索與電腦輔助語言教學研討會」、2004 年「文件探勘技術研討會」、2005 年「網路資訊檢索技術與趨勢研討會」、2006 年「網路探勘技術與趨勢研討會」以及 2007 年「Web 2.0 技術與應用研討會」之後續的年度會議活動。歡迎各界人士踴躍參加。

議程：

時間	講題	主講人
09:00 ~ 09:30	報到	
09:30 ~ 10:30	Annotation Prediction in Social Media Mining	陳信希 教授 (台灣大學資訊工程學系)
10:30 ~ 11:00	休息	
11:00 ~ 12:00	Cloud Computing and The Future of Internet Services	馬維英 副院長 (微軟亞洲研究院)
12:00 ~ 13:30	午餐	
13:30 ~ 14:20	Applying Cloud Computing in Mining Social Network Content	張俊盛 教授 (清華大學資訊工程學系) (中研院國際研究生院 計算語言學學程)
14:20 ~ 15:10	Aspect Summarization from Blogosphere for Social Study	張嘉惠 教授 (中央大學資訊工程學系)
15:10 ~ 15:30	休息	
15:30 ~ 16:20	Link Mining in the Blogosphere	高宏宇 教授 (成功大學資訊工程學系)
16:20 ~ 17:10	Social Network Mining: the past, current, and future.	林守德 教授 (台灣大學資訊工程學系)

網路社群服務計算暨探勘技術研討會

報 名 表

編號：

會議時間：2008 年 12 月 31 日（星期三）
 會議地點：台北市中央研究院資訊所一樓 106 演講廳
 主辦單位：中央研究院資訊所、台灣大學資訊工程學系
 成功大學資訊工程學系、中華民國計算語言學學會

姓 名		單 位	
電 話		E-mail	
聯絡地址			
收據抬頭			
報 名 費	一般人士： <input type="checkbox"/> 會員 400 元 <input type="checkbox"/> 非會員 600 元 學生： <input type="checkbox"/> 會員 200 元 <input type="checkbox"/> 非會員 300 元 (報名費含講義、午餐及茶點)		

1. 學生非會員請附學生證影本
2. 報名及付款截止日 **12/23** 日（現場報名加收 200 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：

信用卡持有人：_____有效期：_____

卡號：_____—_____—_____—_____

卡片末三碼：_____發卡銀行：_____

持卡人簽字：_____ (簽名方式請與信用卡相同)
4. 午餐請備素食
5. 報名表請傳真或 E-mail 至：

Fax. No.: (02)2788-1638, E-mail: jessie@hp.iis.sinica.edu.tw
6. 聯絡資訊：電話：(02)2788-3799*1502, 聯絡人：黃琪 小姐

ACL-IJCNLP 2009 Preliminary Call for Papers

August 2 - 7, 2009

Singapore

<http://www.acl-ijcnlp-2009.org>

For the first time, the flagship conferences of the Association of Computational Linguistics (ACL) and the Asian Federation of Natural Language Processing (AFNLP) --the ACL and IJCNLP -- are jointly organized as a single event. The conference will cover a broad spectrum of technical areas related to natural language and computation. ACL-IJCNLP 2009 will include full papers, short papers, oral presentations, poster presentations, demonstrations, tutorials, and pre- and post-conference workshops. The conference invites the submission of papers on original and unpublished research on all aspects of computational linguistics.

Important Dates:

- Feb 22, 2009 Full paper submissions due;
- Apr 12, 2009 Full paper notification of acceptance;
- Apr 26, 2009 Short paper submissions due;
- May 17, 2009 Camera-ready full papers due;
- May 31, 2009 Short Paper notification of acceptance;
- Jun 7, 2009 Camera-ready short papers due;
- Aug 2-7, 2009 ACL-IJCNLP 2009

Topics of interest:

Topics include, but are not limited to:

- Phonology/Morphology, Tagging and Chunking, Word segmentation
- Syntax and Parsing
- Semantics
- Discourse, Dialogue and Pragmatics
- Language generation
- Summarization
- Machine translation
- Information retrieval
- Information Extraction
- Sentiment Analysis and Opinion Mining
- Question Answering
- Text mining and Natural language processing applications
- Spoken language processing
- Statistical and machine learning methods
- Language resources and Evaluation

如何讓電腦說話? - 論語音合成技術

賴玟杏

國立高雄第一科技大學電腦與通訊系

一、簡介

語音可以說是最自然與方便的人機界面。想像如果電腦能直接以語音來回答問題或提供資訊該有多好?尤其在某些情況，例如開車時，我們的眼睛與手都忙碌於注意路況及開車，如果資訊的提供能以語音的方式將更為安全而方便。對視障同胞而言，如果電腦會說話將可對他們有很大的幫助，也減少很多生活上的不便。而想讓電腦開口說話便要利用到語音合成的技術。當電腦會說話後，能應用的範圍非常廣，例如整合訊息服務系統、語音即時新聞電話查詢、視障資訊快速閱讀、測量台查修指引系統、自動閱報機、有聲編輯系統、氣象及路況報導等。對於需經常更新內容的語音播報及查詢系統，使用語音合成最大的好處是只要有文字輸入，就可即時合成出清晰、自然、流利的語音，可節省大量錄音的人力和時間。

對於所需語句固定的系統，多可以錄音的方式事先將語句及可能選項錄下，將預錄好的文句予以編碼貯存，合成時再將這些儲存資料予以解碼(或根本沒有編解碼過程)，播音時再自動加以簡單連接放音即可。例如目前市面上一般的訂票或掛號系統即屬於這種。這樣的有限文句的系統，雖可利用儲存再播送的技術，但通常受限於文句內容固定，故應用範圍較為有限且不具彈性，但在音質上則較易掌控。但若是語句不固定、不限範圍的語音合成，如即時的新聞播報，則很難以預先錄音的方式處理。針對這樣不限範圍的語音合成系統，我們稱為文句翻語音系統(Text-to-Speech，簡稱TTS)。這樣子的系統不限定合成文句的內容，因此應用範圍較廣，但技術上也較難，早期由於技術比較不成熟，因此要合成一清晰自然的語音，需要花費很多功夫來調整。

文句翻語音系統簡單來說是將一段文句或一篇文章透過語言及語音處理技術將之轉換成自然、流利且清晰的聲音。其工作方式是先將文句轉成相對應之讀音，根據音韻規則，選取適當的合成語音單元，用語音合成器合成出所需的語音。雖然這幾年語音合成的技術進步，但一般的文句翻語音系統所能提供的聲音品質已漸不能滿足人們對高品質合成語音的需求，近年以資料庫為基礎之語音合成方法開始受到重視。雖然所需要儲存的資料較多，但一般而言，在適當的設計下，

因單元相對較長且多，清晰度及自然度都大為提升。但一般文句翻語音系統因其所需資料庫較小，適合客戶(Client)端的語音合成應用，如何改進其音質仍然是熱門的課題。此外，如何合成出帶有喜怒哀樂情緒的語音，或者合成出歌聲都是近來的合成方向。

以下第二節中將介紹文句翻語音系統的架構及各模組工作方式。第三節中將介紹幾個相關的合成方向。最後一節是結論。

二、語音合成系統基本架構

一般的文句翻語音系統，可能包含語言、音韻及合成單元等三個資料庫以及語言分析、音韻預估、合成單元選取、及合成器等四個模組，其架構如圖 1 所示。首先，輸入文章被送入語言處理模組。語言處理模組靠語言資料庫的幫助做斷句、斷詞、構詞、甚至語法、及語意分析等處理，得到正確的讀音。語言處理的結果被送入音韻模組。利用上下文及讀音等資訊，加上音韻資料庫中所提供的韻律規則或由訓練而來的參數，音韻預估模組計算可能的音韻值，如音長、停頓、音量及基頻值等。合成單元選取模組則根據所需的音及預估的韻律值從合成單元資料庫中選取最適當的合成單元。合成器依據音韻預估模組所預估的韻律調整合成單元選取模組所選擇的合成單元，加以適當的連接後輸出合成語音。

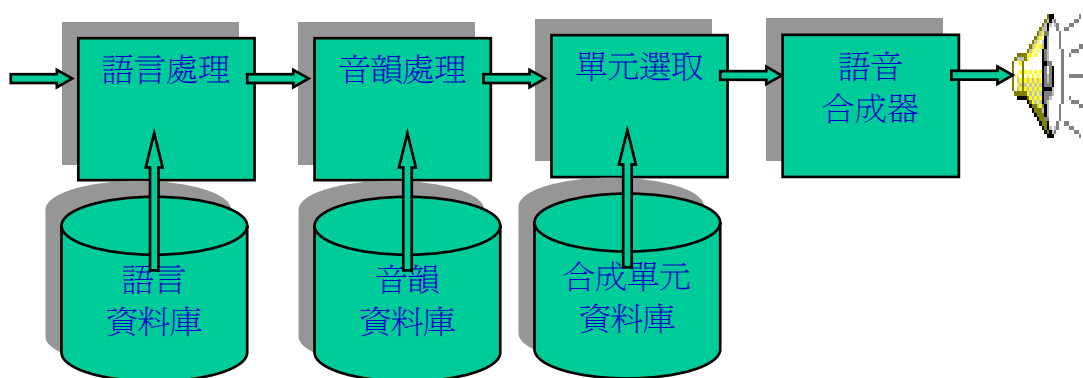


圖 1: 一般語音合成系統基本架構圖

文句翻語音系統所需要的三個資料庫分別介紹如下：

1. 語言資料庫

語言資料庫儲存語言處理分析所需之資料。這些資料可能包括字音表、詞音表、詞類、詞頻、破音字字音表等。

2. 音韻資料庫

音韻資料庫儲存預估音長、音強、基週、停頓時間等音韻資料所需之規則或模型參數。這些模型參數是在訓練階段求得的。

3. 合成單元資料庫

合成單元資料庫儲存語音基本單元，語音基本單元可能是音素、雙音素(diphone)、三音素(triphone)、半音節(demi-syllable)[1]、單音節、詞或者更長的單元。單元的長度可能是固定或不固定的，甚至同一個單元有幾個不同韻律的合成單元版本供選擇。一般而言，所選用之語音基本單元越大，則合成之接點越少，合成語音更容易維持清晰自然，但所需的單元更多，儲存空間也越大。目前許多中文系統採用單音節的字為合成單元，主要原因是中文為一字一音的語言，但是因為缺乏連音的考慮，無法合成出很自然的聲音。

另外，文句翻語音系統的四個模組也介紹如下：

1. 語言處理模組

語言處理模組[2]主要的功能是在語言資料庫的幫助下，分析輸入的文章，予以適當之斷句、斷詞、構詞，甚至分析語法結構及語意，找出該文章每個字的正確讀音，並且提供詞邊界、換氣邊界等相關訊息給韻律處理模組作為韻律預估之用。

在語言處理模組中，首先需經過智慧型前處理。智慧型前處理的目的在於將輸入文句中的符號或數字先轉成恰當的文字。例如在「有 1/3 的人無家可歸」這個句子裡的「1/3」代表三分之一，而不是一月三日。而「1234567」可能念為電話號碼的「一二三四五六七」或者代表金額的「一百二十三萬四千五百六十七」。如果把「有 1/3 的人無家可歸」念為「有一月三日的人無家可歸」或把「電話號碼為 1234567」念為「電話號碼為一百二十三萬四千五百六十七」，一定會讓聽者一頭霧水。

另外，中文的寫法在詞與詞間並沒有分隔符號，因此不管要分析語法結構或決定正確讀音都需要先找到詞。初步斷詞先將字串分成數個包含可能詞彙的大詞段（沒有一個可能詞彙會跨越這些詞段），細步斷詞則在每個詞段裡利用最長匹配法則選取最可能的詞彙組合。但中文詞的變化很多，如果把所有可能的詞都

列入辭典中，不僅使得辭典所佔的記憶容量過於龐大，而且降低在辭典中搜尋詞群的速度，影響到整個文字翻語音系統的效率。所以，為縮短查詢辭典的時間，屬於非規律性字串組合的詞及出現頻率較高的詞應放進辭典中，有規律的部份則在構詞處理中組合成詞。中文主要的三種構詞方式是附加、重疊及複合。決定詞群後，可由詞庫中的標音找到正確讀音，不在詞庫中的可以字轉音表來處理，但仍可能有破音字的問題。

中文大約有百分之十的字有兩個以上的發音。但大部分的可經由辭典及詞頻資訊(選用最常用的發音)來解決，剩下錯誤的機會大約是百分之二到三左右。這個機率數字也許並不顯眼，但是其實只要電腦一唸錯音時，大部份的人都會把焦點注意在這個錯誤。所以一個語音合成系統要能為大眾廣泛接受，必須將錯音的比例盡量降低。但這不是一個容易的問題，例如「我這次得到美國方面鼎力相助，得到美國去出差。」，句中出現兩次「得」的發音都不同，不是簡單的辭典及詞頻方法可解決。要更完整解決破音字的問題仍有賴更完善的語法結構及語意分析，不過目前的技術在語法語意部份仍有待改進。

2. 音韻處理模組

一個語音合成系統要能夠動人，必須具有相當的自然度。而自然的語音有其抑揚頓挫。音韻模組的目的就是從讀音、語法結構、語意等，找出適當的方法來預估合成語句的抑揚頓挫。包括對聲調、音長、音量及停頓等參數的預估。早期音韻處理多採用語言學家經過長期收集觀察而歸納的一些規則，但這種方法曠日費時且較難同時考慮許多因素的影響；近期則偏向利用統計的方法(即資料庫訓練法)。一般的方法是採用音韻模型來代表音韻受各因素影響的情形，先以大量的語料訓練此模型後，使此模型得以代表韻律特性，然後將此訓練後之模型應用於語音合成系統上來預估音韻值。在統計方法上，包括回歸模型(Regression model)、類神經網路[3]、期望最大化(Expectation-Maximization，簡稱EM) [4] [5] 等方法都有不錯的成效。

統計模型可以幫助我們定量地決定每個因素影響音韻值的程度。而到底有哪些因素會影響音韻呢？對於聲調、音長、音量及停頓而言，不同的因素都各有不同程度的影響。一般可能影響音韻的因素包括：

- 目前的及相鄰的聲調
- 目前的及相鄰的音素、子音、母音、音節等
- 詞類、片語種類(如名詞片語、動詞片語)
- 在詞、片語、換氣群及句中的位置
- 詞、片語、換氣群及句的長度
- 語者

- 標點符號
- 語意、情緒等

3. 合成單元的選擇

合成單元選取技術是從資料庫中選擇合適的合成單元來連接產生需要的語句，是文字翻語音系統中經常使用來提高聲音品質的方法。單元選取技術的引進通常可以相當地提升合成聲音的品質。此技術主要是從語言學特性或音韻特性上找到一些選擇的標準以選出最符合目標值的最恰當單元。如果單元語料庫夠大，含有足夠多不同上下文及韻律變化的組合，則調整單元就變得不需要或只要做小幅的調整。而調整單元通常是造成聲音品質下降的重要原因。因此這種方式合成出來的語音通常會比單一單元的合成語音更為清晰自然。合成單元選擇的方法很多，例如樹狀結構(Tree-based) [6]、隱藏式馬可夫模型(HMM-based) [7]的方法。有些研究利用定義成本(cost)的方式來實現選取的準則，例如使用結構句法成本(Structural Syntactic Cost) [8]。

4. 合成器

人類的發聲器官就像樂器一樣，可以發出很多種聲音，但也有許多先天的限制。空氣由肺部送出，如果是濁音，聲帶會震動，氣流在發聲腔道內產生共振現象，然後從嘴部釋放出來，產生聲波，傳到我們的耳朵。如果是清音，激發源則為亂流的雜音型態。而靠著不同的發音位置及方式可發出不同的聲音。但若以簡單的線性發聲管模型來模擬人類的發音，仍嫌不足。因而有許多的分析合成模型或合成器被提出。

合成的方法除了用一組不同口徑相連的發聲管來模擬聲道外，另有最早貝爾實驗室(Bell Labs)的 Homer Dudley 在 1939 年所提出的 vocoder。其他還有共振峰(formant)合成器、時域上的基週同步疊加法(Pitch-Synchronous Overlap Add，簡稱 PSOLA)、及之後的弦波模型(sinusoidal modeling)等方法。

以單脈衝／噪音激發之線性預估編碼為例，此種基於口腔模型的合成器可以隨意地調整其基週週期，然而合成語音品質受限於單脈衝激發(single pulse excitation)[9]的聲帶模型，語音較粗糙。而品質比較好一點的是以多脈衝激發(multipulse excitation) [10]或是碼本激發(codebook excitation) [11]的方式。

之後受到歡迎的是所謂的波形合成器，此類合成器在時域直接調整語音。一般而言，若是調整音韻的範圍不大，原始語料的音質不會受破壞，因此合成語音的品質會比線性預估編碼合成器好。基週同步疊加法即是一套在時域直接修改語音的演算法。這是比較簡單又能得到不錯音質的方法，因而受到歡迎。

此外，McAulay 和 Quatieri 採用弦波模型[12]在分析合成上。George 和 Smith 也提出了 Analysis-by-Synthesis/Overlap-Add (ABS/OLA 方法)[13]。而後又有 Harmonics plus Noise Model (HNM) [14]的提出。其他還有許多改進的方法。然而距離真實重製人類的聲音仍有相當的研究空間。

三、其他相關的語音合成方向

除了以上所介紹的一般的文字翻語音系統，語音合成仍有很多的研究方向。以下列出其中幾項。

1. 以大量語料庫為基礎的(Corpus-based)語音合成

傳統語音合成的優點有：資料庫所需記憶體較小，系統可攜性較高。而缺點為合成語音自然度較難提升，因為要將有限的語音單元合成為自然語音，語音單元必須經過轉換調整，經過大幅調整後的單元不容易維持音質，另外因為選擇太少，也難達到自然。

於是更龐大的音韻資料庫被用來訓練音韻參數模型，更大量更多樣化的語音合成單元被用在合成中。配合電腦硬體之運算速度越來越快，儲存容量越來越大，價格卻越來越低廉，於是設計一個可容納自然語言變化的龐大資料庫變得可能，以大量語料庫為基礎的語音合成系統[15]也就因此應運而生。因為語音單元已包括許多變化，在單元相接時，不需對合成單元做太大更動就能產生接近自然原音的合成語音。

2. 情緒語音合成

語音一向是一種最方便的人機介面。電腦語音在過去幾年裡進步不少，變得更加易懂、易聽。但跟真人相比，聲音裡面還是缺少了一些元素，那就是情緒。人類情感的表達方式很多，透過語音的情緒表達是溝通很重要的一部分。如果能在電腦合成聲音中再賦予情緒，就能讓合成語音更真實，也能溝通表達得更好。目前已經有相當多學者及機構投入情緒語音合成的研究[16] [17]。

3. 歌聲合成

唱歌跟說話一樣都是人類溝通及表達的自然方式。歌聲合成的應用包括作曲、娛樂、教學等領域。但是歌聲合成雖與語音合成一樣是處理人類發聲的問題，卻有著跟語音合成不一樣的困難。在歌聲合成中，作曲者已經在譜中提供了韻律這方面的訊息，但很可惜地，在合成時，若只依照譜上的標準韻律值來合成，並無法產生自然的歌聲。當我們在唱一首歌時，我們會試著去表達歌詞，而職業的歌者更賦予歌聲許多藝術成分。這使得歌聲的變化比起說話時更為動態而複雜，且變化的範圍更大。例如顫音是唱歌時特有的現象。又例如女性古典歌手會將第

一個共振峰移至接近基週的位置，以產生更有力的聲音，使其可在其他樂器或歌聲中更容易被聽見。另外，濁音(voiced)的出現比率與說話時也有很大的不同。歌唱時濁音出現的比率大約是百分之九十五[18]。相對地，英文說話語音中濁音與清音(unvoiced)及無聲(silent)的比率約為百分之六十比二十五比十五。在從事歌聲合成的研究時，這些因素都要列入考慮，才能開發出高品質的歌聲合成系統。

四、結論

雖然目前的合成語音跟真人聲音間還有距離，但讓電腦說話已不是達不到的夢想。其實在許多播報系統上，語音合成都已進實用的階段。終有一天，這項科技會用在所有家電、影片、電玩、動畫、機器人或所有我們能想到或想不到的應用領域，發出栩栩如生的聲音。

參考文獻

- [1] C.S. Liu, W.J. Wang, S.M. Yu and H.C. Wang, "Mandarin Speech Synthesis by the Unit of Coarticulatory Demi-syllable," *International Conference on Spoken Language Processing*, Kobe, Japan, Nov. 1990.
- [2] Chun-Hsiao Lee, Eng-Fong Huang and Chun-Jen Lee, "A Linguistic Processing Module in the Chinese Text-to-Speech System," *Bimonthly Journal of Telecommunication Research*, Vol. 23, no. 1, Feb. 1993.
- [3] S. H. Chen, S. H. Hwang and Y. R. Wang, "An RNN-based prosodic information synthesizer for Mandarin text-to-speech," *IEEE Trans. Speech and Audio processing*, vol. 6, no.3, pp.226-239, 1998.
- [4] Sin-Horng Chen, Wen-Hsing Lai, Yih-Ru Wang, "A statistics-based pitch contour model for Mandarin speech," *Journal of the Acoustical Society of America*, 117 (2), pp. 908-925, February 2005.
- [5] Sin-Horng Chen, Wen-Hsing Lai, Yih-Ru Wang, "A New Duration Modeling Approach for Mandarin Speech," *IEEE Trans. Speech and Audio processing*, vol. 11, no.4, pp.308-320, July 2003.
- [6] W.J.Wang, N.Campbell, N.Iwahashi and Y.Sagisaka, "Tree-based Unit Selection for English Speech Synthesis," *ICASSP*, pp. 191-194, 1993.
- [7] Zhen-Hua Ling, Ren-Hua Wang, "HMM-Based Hierarchical Unit Selection Combining Kullback-Leibler Divergence with Likelihood Criterion," *ICASSP*, Volume 4, 15-20 April, 2007, Page(s):IV-1245 - IV-1248.

- [8] Chung-Hsien Wu, Chi-Chun Hsia, Jiun-Fu Chen, and Jhing-Fa Wang, "Variable-Length Unit Selection in TTS Using Structural Syntactic Cost," *IEEE Transactions on Speech and Audio Processing*, VOL. 15, NO. 4, pp. 1227- 1235, MAY 2007.
- [9] C.S. Liu, M.F. Tsai, Y.H. Hsyu, C.C. Lu and S.M. Yu, "A Chinese Text-to-speech Based on LPC Synthesizer," *Telecomm. Labs. Technical Journal*, Taiwan, Vol.19 No.3 pp.269-285, Sept. 1989.
- [10] C.S. Liu, G.H. Ju, W.J. Wang, H.C. Wang and W.H. Lai, "A New Speech Synthesizer for Text-to-Speech System Using Multipulse Excitation with Pitch Predictor," *Proceedings of ICPCOL*, Taiwan, Aug. 1991.
- [11] Shu-fung Yang, Gwo-hwa Ju, Chun-hsiao Lee, Wen-hsing Lai and Chi-shi Liu, "A Chinese Text-to-Speech System Based on Modified CELP Speech Codec," *IEEE International Workshop on Intelligent Signal Processing and Communication Systems*, 1993, Japan.
- [12] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 4, pp. 744–745, 1986.
- [13] Bryan George and Mark J. T. Smith, "Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 389-406, September 1997.
- [14] Yannis Stylianou, "Applying the Harmonic Plus Noise Model in Concatenative Speech Synthesis," *IEEE Transactions on Speech and Audio Processing*, 2001, vol. 9, pp. 21–29.
- [15] Fu-Chiang Chou; Chiu-Yu Tseng, "Corpus-based Mandarin speech synthesis with contextual syllabic units based on phonetic properties," *ICASSP*, Volume 2, 12-15 May 1998 Page(s):893 - 896 vol.2.
- [16] Akemi Iida, Nick Campbell, Fumito Higuchi, Michiaki Yasumura, "A corpus-based speech synthesis system with emotion," *Speech Communication*, 40 (2003) 161–187.
- [17] Dan-ning Jiang¹, Wei Zhang², Li-qin Shen², Lian-hong Cai, "Prosody Analysis and Modeling for Emotional Speech Synthesis," *ICASSP* 1995.
- [18] Perry Raymond, "Identification of Control Parameters in an Articulatory Vocal Tract Model, with Applications to the Synthesis of Singing," *Ph.D. thesis*, Department of Electrical Engineering, Stanford University, September 1991.