

本期要目

- | | |
|---|--------|
| 壹、ROCLING 2010 CALL FOR PAPERS | 第二頁 |
| 貳、CFP- IEEE Transactions on Affective Computing | 第三頁 |
| 參、專文-現階段語音文件摘要研究之簡介(林士翔、陳柏琳) | 第四~十六頁 |

獎助學生出席國際會議

獎助會議：

- | | |
|--------------|-----------|
| 1. COLING | 2. ACL |
| 3. ACM SIGIR | 4. ICASSP |

獎助說明：

- 申請人須同時具備下列資格：
 - 被接受論文之第一作者(指導教授不計)。
 - 本會會員。
 - 投稿時為國內在學學生。
- 獎助金額：由審查委員會依地區別及論文等級審定獎助金額，每名獎助金額上限為美金1,000元。
- 獎助名額：每個會議獎助一~二名。

申請辦法：

- 申請期限：論文被接受發佈日起兩週內提出。
- 申請手續：申請人需將論文接受函、審查意見、學生證、論文全文及申請書等相關資料郵寄至本會秘書處。(申請書請至<http://www.aclclp.org.tw/doc/fundreg.htm> 下載)

受獎助人義務：

- 出席會議發表論文。
- 論文全文必須以書面同意投稿至本會期刊。
- 代學會攜去宣傳品及帶回相關資料。

IJCLCLP - Call for Papers

Special Issue on Web 2.0 and Social Computing

With the popularity of Web 2.0 in recent years, there are an increasing number of applications targeting at not only content but people and their social connections. People interact with each others throughout social networking media such as blogs, Wikis, Flickr, Twitter, and Del.icio.us. Consequently, technologies developed for natural language processing have been widely applied on analyzing and extracting social media. This special issue brings together the computational linguistics with Web 2.0 and social computing. Through soliciting the state-of-the-art research methods and results on the relevant topics, this special issue aims at discovering synergies among the computational linguistics, Web 2.0 and social computing in order to identify new ways of understanding this emerging field.

Areas of interest include, but are not limited to:

- Natural Language Applications on Web 2.0 and Social Networks
- Social Network Analysis
- Search and Mining on Web 2.0 and Social Media

Schedule

Submission deadline:	July 1, 2010
Notification of acceptance:	September 1, 2010
Final manuscript due:	November 1, 2010
Tentative publication date:	December 15, 2010

For more details, please visit our website:
<http://www.aclclp.org.tw/journal/cfp.php>

Call for Papers

Conference on Computational Linguistics and Speech Processing

第二十二屆自然語言與語音處理研討會, ROCLING XXII

National Chi Nan University, Nantou, Taiwan

1- 2 September 2010

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

Important Dates:

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



The new *IEEE Transactions on Affective Computing* seeks original manuscripts for publication. This new online only journal will publish cross disciplinary and international archival results of research on the design of systems that can recognize, interpret, and simulate human emotions and related affective phenomena. The journal will publish original research on the principles and theories explaining why and how affective factors condition interaction between humans and technology, on how affective sensing and simulation techniques can inform our understanding of human affective processes, and on the design, implementation and evaluation of systems that carefully consider affect among the factors that influence their usability. Surveys of existing work will be considered for publication when they propose a new viewpoint on the history and the perspective on this domain. The journal covers but is not limited to the following topics:

Sensing & Analysis

- Algorithms and features for the recognition of affective state from face and body gestures
- Analysis of text and spoken language for emotion recognition
- Analysis of prosody and voice quality of affective speech
- Recognition of auditory and visual affect bursts
- Recognition of affective state from central (e.g. fMRI, EEG) and peripheral (e.g. GSR) physiological measures
- Methods for multi-modal recognition of affective state
- Recognition of group emotion
- Methods of data collection with respect to psychological issues as mood induction and elicitation or technical methodology as motion capturing
- Tools and methods of annotation for provision of emotional corpora

(Cyber)Psychology & Behavior

- Clarification of concepts related to 'affective computing' (e.g., emotion, mood, personality, attitude) in ways that facilitate their use in computing.
- Computational models of human emotion processes (e.g., decision-making models that account for the influence of emotion; predictive models of user emotional state)
- Studies on cross-cultural, group and cross-language differences in emotional expression
- Contributions to standards and markup language for affective computing

Behavior Generation & User Interaction

- Computational models of visual, acoustic and textual emotional expression for synthetic and robotic agents

- Models of verbal and nonverbal expression of various forms of affect that facilitate machine implementation
- Methods to adapt interaction with technology to the affective state of users
- Computational methods for influencing the emotional state of people
- New methods for defining and evaluating the usability of affective systems and the role of affect in usability
- Methods of emotional profiling and adaptation in mid- to long-term interaction
- Application of affective computing including education, health care, entertainment, customer service, design, vehicle operation, social agents/robotics, affective ambient intelligence, customer experience measurement, multimedia retrieval, surveillance systems, biometrics, music retrieval and generation

Editor in Chief: Jonathan Gratch, University of Southern California. To submit a paper, go to <https://mc.manuscriptcentral.com/taffc-cs>.



Published in cooperation with: IEEE Robotics and Automation Society, IEEE Signal Processing Society, IEEE Society on Social Implications of Technology, IEEE Consumer Electronics Society, IEEE Circuits and Systems Society

現階段語音文件摘要研究之簡介

林士翔、陳柏琳

國立台灣師範大學資訊工程系

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

一、前言

在現今資訊科技與網際網路快速蓬勃發展的時代，大量的文字或多媒體影音資訊被快速地傳遞與分享於全球各地，資訊超載(Information Overload)問題也因此隨之產生。如何能讓人們快速地、有效率地瀏覽與日俱增的文字資訊或多媒體影音資訊，已是一個刻不容緩的研究課題。而在眾多的研究方法中，文件摘要(Document Summarization)被視為是一項不可或缺的關鍵技術[1-4]。

文件摘要旨在於擷取單一文件(Single-Document)或多文件(Multi-Document)中內含的重要語意與主題資訊，並以適當的方式呈現，讓使用者可以快速地瀏覽與理解文件中的重要資訊。因此，使用者不需花費大量的時間與力氣逐一地審視或過濾文件內容，便可以快速獲得其所需的資訊。另一方面，文件摘要技術對於如何能自動地、有效率地處理具時序性的多媒體影音內容，諸如電視及廣播新聞、語音郵件、會議及演講錄音等，更是顯得非常重要[3]。其原因在於對於動輒數分鐘到數小時的多媒體影音，使用者通常無法如瀏覽文字文件般的任意(或是隨機)跳躍地瀏覽文件中的任何一個部份，而必須耐心地從頭到尾閱讀或聆聽整份多媒體影音內容，才能理解其中所描述的語意與主題。雖然對於含有語音訊號的多媒體影音，我們可透過自動語音辨識(Automatic Speech Recognition, ASR)的技術將其轉換成易於瀏覽的文字內容，再經由文字文件摘要(Text Document Summarization)技術的處理，達到摘要多媒體影音或者其他語音文件(Spoken Documents)的目的。但就現階段語音辨識技術的發展而言，語音文件經語音辨識自動產生的文字轉寫結果，仍存有一定程度的辨識錯誤，並且缺乏像章節與標點符號等結構資訊；再加上語音文件本身所包含的許多口語助詞、遲疑、重覆等內容，使得語音文件摘要(Spoken Document Summarization)技術的發展面臨更多的挑戰。所以，近年來語音文件摘要技術的發展在語音及自然語言處理領域已是一個愈來愈受重視的研究課題[2-3]。

由於目前仍有許多語音文件摘要研究所使用的摘要技術是源自於傳統文字文件摘要研究的成果，因此本文將先簡略地回顧過去在文字文件摘要上一些重要的相關研究與成果；然後，闡述當前語音文件摘要研究的發展現況，並介紹數個文件摘要效能的評估方式；最後是結論與對於語音文件摘要研究的未來展望。

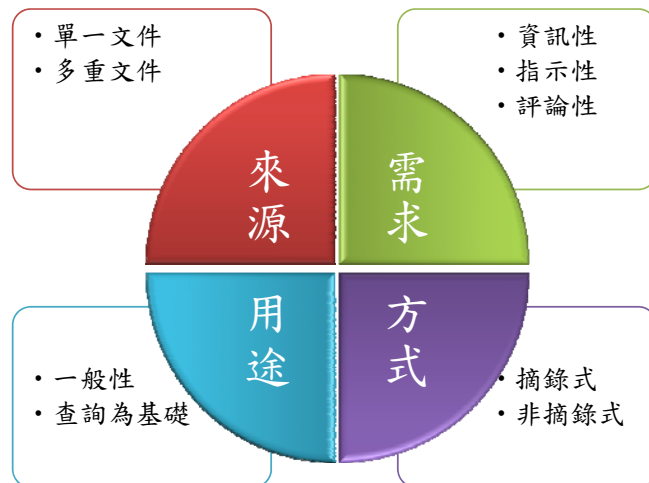


圖 1. 自動摘要技術分類圖

二、文字文件摘要技術

一般而言，文件摘要技術可從許多不同構面進行探討，包括來源、需求、用途及方式，分類圖如圖 1 所示，下列將簡述各個不同構面的差異[1]：

1. 來源：根據文件來源，可以分為單一文件摘要與多文件摘要；而在多文件摘要中，尤其需額外考慮文件間彼此的資訊重複(Redundancy)性[5]與文件所描述事件的發生先後順序(Causality)[6]。
2. 需求：依據使用者需求不同，摘要內容可以是具有資訊性(Informative)，用來表達文件本身所含的重要語意及主題資訊；或者具指示性(Indicative)，可提供如文件分類等文件處理之依據；亦或具評論性(Critical)，分別將文件正面及反面觀點(Positive and Negative Sentiments)[7]呈現出來。
3. 方式：可概分為二大類，摘錄式(Extractive)摘要與非摘錄式(Abstractive)摘要(或重寫摘要)。前者主要是依特定摘要比例，從原文件中選出重要的文句、段落或章節，經適當的修飾與串接而組成摘要；而後者則是根據文件內容自動地重新產生一段簡短扼要的摘要來代表文件的語意及主題資訊，其中所使用之詞彙或慣用語不一定是全然地來自於原始文件，是一種較為貼近人們日常撰寫摘要的方式。
4. 用途：依摘要用途可分為一般性(Generic)摘要與以查詢為基礎(Query-based)的摘要，前者呈現文件全面性的主題，摘要內容以涵蓋整篇文件所有重要主題為主；後者則根據使用者或特定的查詢來呈現文件中與查詢相關的摘要資訊。

然而因為非摘錄式摘要需要較多複雜的自然語言處理(Natural Language Processing, NLP)技術，如資訊擷取(Information Extraction)及對話理解(Discourse Understanding)及自然語言生成(Natural Language Generation)等技術[8-9]，因此近年來研究主要還是以摘錄式摘要為主。自動摘要技術發展概況如圖 2 所示，以下將就摘錄式摘要的研究與進展作簡短回顧。

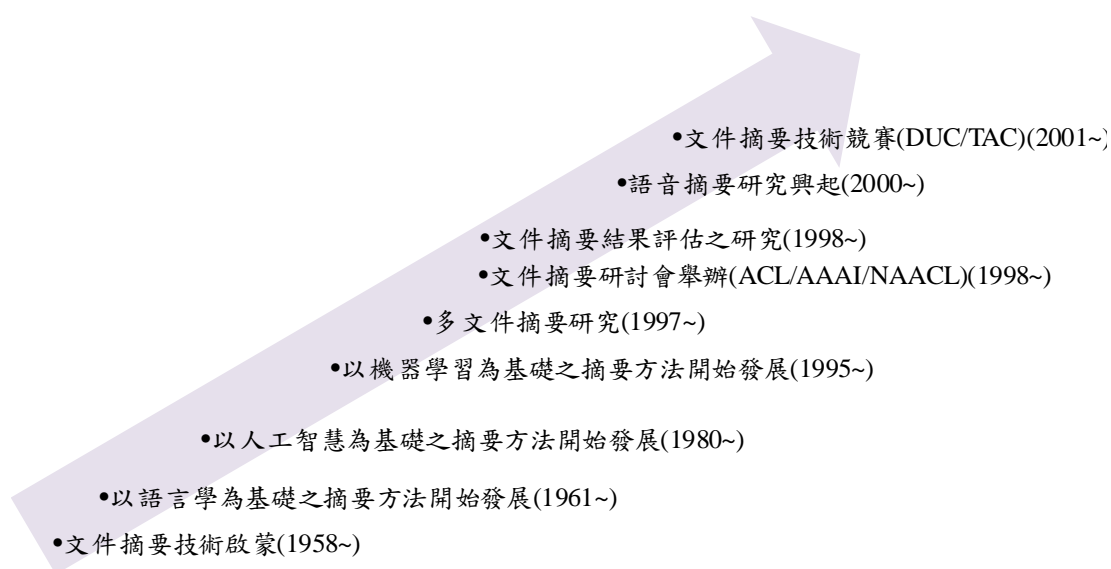


圖 2. 自動摘要技術發展概況

傳統文字文件摘要技術發展最早可追溯至 1950 年代末期，IBM 學者 Luhn 提出以使用詞頻(Frequency)來評量詞的重要性與計算文件中每個語句的顯著性(Significance Factor)[10]，再將語句依其顯著分數進行排序(由高至低)，最後根據特定摘要比例進行摘錄式摘要的產生。此研究不僅自此開啓了自動文件摘要研究的大門，同時亦奠定了許多日後研究可以依循的準則，例如每個詞彙可以進行詞幹分析(Stemming)將其還原成詞根(Root Form)、移除停用詞(Stop Word)的影響，及計算實詞(Content Word)的重要性等。隨後，學者 Baxendale 在審視 200 篇科技文章後，發現有 85%的重要語句會出現在文章中的第一段、以及 7%的重要語句會出現在最後一段[11]。因此，提出語句的在文章中的位置(Position)資訊是在進行文件摘要的重要語句選取時的一項關鍵線索。常常被拿來定義使用。再者，學者 Edmundson[12]除了使用詞頻與位置資訊外，並提出可以額外地使用線索慣用語(Cue Phrases)與詞彙出現在文件不同部份(Skeleton)(如標題、前言、結論等位置)的頻率特徵，再結合統計式方法，對上述這些種特徵作線性組合以計算語句重要性。經過上述三個先探研究(Pilot Studies)後，文件摘要便逐漸成為自然語言處理的一項重要的研究課題。以下，本文將過去研究所陸續發展出的文件摘要方法概略地區分為二大類 [1]：

1. 以語段為基礎(Discourse-based Approach)之摘要方法

此類方法通常是利用自然語言分析(Natural Language Analysis)技術對文章結構進行剖析，接著再根據文法結構(Grammar Structure)與語言機制(Linguistic Devices)，如

首語重複(Anaphora)、省略(Ellipsis)、結合(Conjunction)，或同義詞(Synonymy)、上義詞(Hypernym)等詞彙關係(Lexical Relation)決定不同語段的凝聚(Cohesion)關係，以進行自動摘要。此類方法較著名的相關研究包括使用詞彙鏈(Lexical Chain) [13]、宏觀語段結構(Discourse Macro Structure)[14]、修辭結構(Rhetorical Structure)[15-17]等。

2. 以機器學習為基礎(Machine-learning-based Approach)之摘要方法

此類的研究可追溯至於 1990 年代，因機器學習(Machine Learning)技術逐漸被介紹至自然語言處理而興起。一般來說可分為兩個面向：非監督式(Unsupervised)摘要模型與監督式(Supervised)摘要模型的使用。非監督式模型通常產生單一種摘要特徵供語句排序使用，例如語句與文章相關性[18]、語句所形成的語言模型生成文件之機率等[19]、語句間關係性[20-21]、語句中詞彙在機率式潛藏主題空間(Probabilistic Latent Topic Space)的重要性[22]、或語句與文件在機率式潛藏主題空間的模型距離[23-25]等，這些作法基本上在建立摘要模型時並不需參照或使用任何人工事先標記過的語料庫。另一方面，監督式模型則通常需要有人工事先標記過的語料庫供模型訓練使用，亦即會使用到一組訓練文件集以及其中每篇訓練文件的每一語句是否屬於摘要語句(Summary Sentence)的資訊來做為模型訓練之基礎。通常，監督式模型可同時結合多種摘要特徵來表示每一語句(其中的每一種特徵可以是上述以語段為基礎摘要方法或是非監督式摘要模型針對此語句所輸出的分數或機率值)，這些特徵合起來所形成的特徵向量可來用來做為監督式摘要模型判斷此語句是否屬於摘要語句的依據[26]。此類方法較著名的相關研究包括簡單貝氏分類器(Naïve-Bayes Classifier)[27]、高斯混合模型(Gaussian Mixture Model, GMM)[28]、隱藏式馬可夫模型(Hidden Markov Model, HMM)[29]、支援向量機(Support Vector Machine, SVM)[30]、條件隨機場域(Conditional Random Fields, CRF)[31-32]等。

三、語音文件摘要技術

目前多媒體影音已經佔據了網際網路一半以上的流通量，雖然透過視覺線索可以對多媒體影音內涵進行瀏覽與搜尋，但其所含的語音文件卻能提供更豐富的語意描述，諸如文字轉寫、語者、情感、說話時場景等資訊。因此，文件摘要研究在最近十年來已將觸角延伸至語音文件摘要，常使用的研究題材包括了廣播新聞(Broadcast News)[19, 33-34]、語音郵件(Voice Mails)[35-36]、會議錄音(Meeting Recordings) [28, 31, 37]、演講紀錄(Lecture Recordings)[38]等。早期較直覺式的語音文件摘要作法是直接將語音文件利用自動語音辨識技術，將語音訊號轉換為文字轉寫，然後採取與文字文件摘要相同的技術以得到語音文件的自動摘要。然而，此舉卻突顯出在從事語音文件摘要時會面臨傳統文字文件摘要所未曾遭遇問題，例如語句中往往會帶有語音辨識錯誤、遲疑、重覆或無意義的語助詞等錯誤或不必要資訊。再者，因為語音文件的文字轉寫通常是以純文字(Plain Text)的方式呈現，缺乏像章節、標點符號、字體大小、句子邊界等結構資訊，以致於傳統文字文件摘要方法並無法正確地摘要出語音文件中重要文句的問題[2]。此外，語音文

件擁有許多文字文件所沒有的資訊，諸如語調(Intonation)、音高(Pitch)、能量(Energy)、停頓時間(Pause Duration)等聲韻特徵(Prosodic Features)，或是語者、情感、說話時場景等資訊，都是從事語音文件摘要時可以善加利用的額外語句特徵來源。因此，如何能善用這些語音訊號本身所獨具的特性來提昇語音文件摘要的成效，已成為當前語音文件摘要研究的新興課題之一[2, 39]。另一方面，就自動語音文件摘要而言，摘錄式摘要可以保留原始語者在當時說話的語調及情感等資訊，若藉由直接播放摘要語句所對應的語音訊號不但可以呈現這些資訊，亦可免去語音辨識錯誤所造成的影響[40]。因此，近年來有關語音文件摘要的研究，大部分是以摘錄式摘要為主，並且可以歸納分為四個主要流程，如圖 3 所示，下列將介紹每個階段的主要研究方向：



圖 3. 摘錄式語音摘要技術流程圖

1. 資料前處理(Pre-processing)

在資料前處理階段，主要研究方向又可細分數個研究內容，如語音偵測、語者辨識、語音辨識或是移除口語對話所產生的負面影響等。然而相較其他語音研究，最主要的不同點在於如何選定摘錄式摘要的摘錄單位，通常是語句或者更大的單位像是段落或章節等。但如同前面所述，語音文件通常沒有正確的結構資訊。因此，如何定義出語音文件中摘要單位的邊界，便是從事語音文件摘要一大挑戰。因為邊界定義的好壞，可能會直接影響到摘要結果的閱讀品質或語音播放流暢度。目前常見的作法有使用語音訊號停頓(Pause)資訊來做為摘要單位切割的參考依據；在藉由訂定閾值(Threshold)後，可在當語音停頓時間超過該閾值時標示為一個可能摘要單位的邊界。此外，亦有學者利用語言模型(Language Model)與韻律模型(Prosodic Model)等所計算出的特徵值，結合分類器像是隱藏式馬可夫模型(HMM)、最大熵值法(Maximum Entropy, ME)或條件隨機場域(CRF)等進行邊界偵測[41]；近年來更有研究學者使用語調片語(Intonational Phrase)資訊來進行摘要單位切割[42]。

2. 特徵分析與擷取(Feature Analysis and Extraction)

因為監督式摘要模型往往需使用一組事先已定義好的特徵來描述每一語句或摘要單位。例如，每一語句可以使用 L 個維度的特徵向量來表示。在過去幾年，文獻上已經有許多的不同性質的特徵被成功地發展並應用於語音文件摘要，這其中最為被廣泛地使用的特徵大致上可歸納成下列四類[26, 31, 38-39]：

- (1) 詞彙特徵(Lexical Features)：類專有名詞(Named Entities)個數、停用詞個數、實詞個數、語言模型分數、詞頻、反文件頻率(Inverse Document Frequency, IDF)、詞性標記(Part-of-speech, POS)等。
- (2) 聲學特徵(Acoustic Features)：包括了像語音訊號的強度(Intensity)、音高(Pitch)、共振峰(Formant)，或者詞彙及語句發聲的持續時間(Duration)、說話速率(Speaking Rate)、語音辨識結果的信心度分數(Confidence Score)等。
- (3) 相關度特徵(Relevance Features)：通常是來自於不同的非監督式模型所產生之摘要特徵，如語句與文件在向量空間的相近程度(Similarity)、語句間所形成的向心性(Centrality)[20-21]、語句所建立的語言模型生成文件之機率[19, 26]等。
- (4) 結構特徵(Structural Features)：語句位置資訊、語者角色(Speaker Role)、轉換點(Turn)、段落類型(Segmentation Type)等。

3. 摘要模型(Summarization Model)

在語音文件摘要模型的研究上，除了有研究直接採用與傳統自動文字文件摘要相同的方法以外，亦有許多研究嘗試建立新的摘要模型，期望能更適用於語音文件摘要。例如，利用一線性組合來彙整不同的特徵分數，以求算每一語句的重要性，其中的特徵分數包括詞的語言學分數(Linguistic Score)、重要性分數(Significance Score)及語音辨識結果的信心度分數(Confidence Score)、及語意相依度分數(Semantic Dependency Score)等[40, 43]；或是直接以語音訊號的頻譜特徵向量來做為觀測向量(Observation Vectors)而沒有使用任何詞彙特徵，以避免受到辨識錯誤的影響[44]；亦有研究使用語音辨識所產生的 N -最佳序列(N -Best List)或是詞圖(Word Lattice)等多重轉寫結果來降低辨識錯誤的影響[25]；或是使用翻譯模型(Translation Model)移除語音文件中口語不流暢性(Disfluency)問題[45]。此外，值得一提的是，語音文件的摘要語句常含有多餘或較不重要的詞彙(如綴詞、語助詞等)、以及語音文件辨識錯誤所產生的不正確詞彙，使得基於語音辨識所產生的文字轉寫而獲得的摘要內容往往不夠精簡或者包含雜訊。因此，也有學者進一步提出摘要語句簡化與壓縮的方法[40, 43, 46-48]，以產生更具品質的摘要內容。

4. 資料後處理(Post-processing)

語音文件的摘要內容可以語音辨識產生的文字轉寫或原始語音訊號來呈現。但如同本文前面所敘述的，語音辨識可能會產生錯誤的轉寫文字，若直接使用轉寫文字做為摘要的呈現方式，不僅語者的語調與情感等資訊會遺失，同時也會因語音辨識錯誤的影響而傳遞錯誤的語意或主題資訊。於是，另外一種作法便是直接由原始語音

檔案中根據自動摘要結果切割出相對應的語音訊號內容，串接而成爲輸出的摘要，其優點是保留了原始語音檔案中的語者、語調、情感等資訊，並且不會包含有辨識錯誤資訊，但缺點是有時語音段落的串接部分會有不流暢的情形發生[40, 43]。

四、文件摘要的效能評估

直至今日，文件摘要的效能評估方式仍然是百家爭鳴，尚未真正有一套堪稱完整、且各方都可以信服的方法被提出。目前被廣泛使用的評估方式，大致上可從主觀(Subjective)評估與客觀(Objective)評估來分類。主觀評估基本上是以人的主觀判斷與評估做爲文件摘要的效能優劣判斷；而客觀評估則通常是以計算自動摘要與參考摘要彼此間的相似或相關程度來做爲評估效能。一般而言，參考摘要通常是藉由人工對於文件的語句作“摘要”與“非摘要”標記、或根據文件語意以及主題資訊重寫一段文字而形成摘要。因此，在參考摘要的產生過程中必然會隱含了個人的主觀意見。爲了避免參考摘要之間的差異太大，通常對於同一份被摘要文件而言，會同時準備數份由不同人所人工建立的摘要做爲評估時的參考摘要，藉由自動摘要與不同參考摘要間的主觀評估結果或是客觀評估結果的平均來做爲文件摘要的效能評估值。

1. 主觀評估

主觀評估是由數位評估者對自動摘要結果進行評估，評估的重點包括摘要的可讀性(Readability)與流暢性(Fluency)。評估者通常會跟據下列評估準則：文法正確性(Grammaticality)、非冗餘性(Non-redundancy)、指示清晰性(Referential Clarity)、聚焦性(Focus)、結構性(Structure)與一致性(Coherence)等因素，然後藉由給予不同等級或分數來評定自動摘要的效能[49]。

2. 客觀評估

較常見的客觀評估有：

- (1) 摘要準確率(Summarization Accuracy)：將自動摘要與參考摘要分別以詞串(Word Sequence)來表示，然後將自動摘要與參考摘要對齊(Alignment)，最後以參考摘要的長度減掉自動摘要還原到參考摘要時所需最少的代換(Substitution)、插入(Insertion)及刪除(Deletion)動作個數和，並將之除以參考摘要的長度後所獲得的數值視爲自動摘要的準確率[50]。值得一提的是，在[51]亦有類似的做法，但是不同的地方是[51]先將多份的參考摘要合併成一個詞圖，然後由詞圖中選出一條與自動摘要最相近的詞串，然後計算二者之間的準確率。
- (2) 文句的召回率(Recall)／精確率(Precision)：計算被正確地摘要出的語句佔參考摘要語句總數的比例(召回率)，以及正確地摘要出的語句佔被自動摘要語句總數的比例(精確率)[51]。
- (3) 餘弦評估(Cosine Measure)[52]：計算自動摘要結果與參考摘要分別的向量表示式間的餘弦相似度(Cosine Similarity)來做爲摘要正確率。

- (4) ROUGE 評估方法(Recall-Oriented Understudy for Gisting Evaluation)：其為一種召回率導向(Recall-Oriented)自動摘要評估方法，作法是計算自動摘要與參考摘要之間的重疊單位元次數佔參考摘要長度(單位元總個數)的比例，其中所使用單位元可為 n-連詞(n-gram)、最長共同子序列(Longest Common Subsequence)或詞對(Word Pairs)[53]等。一般而言，單連詞(ROUGE-1)可用來評估自動摘要結果的資訊性(Informativeness)；雙連詞(ROUGE-2)用來審視自動摘要結果的流暢性(Fluency)；而最長共同子序列(ROUGE-L)則可用來衡量自動摘要結果的內容性與文法性(Grammaticality)。

除了上述介紹的評估方法外，目前還有一些其他的文件摘要評估方法有被使用，例如 Kappa 一致性係數[54]、基本元素(Basic Elements)評估[55]、摘要內容金字塔(Pyramid)評估[56]等。

五、結論與未來展望

文字文件摘要與語音文件摘要技術已被發展多年。相較於文字文件摘要，語音文件摘要面對更多的問題與挑戰。經由本文所作的文獻回顧，我們可以清楚地發現現階段語音文件摘要研究仍然存在有許多值得努力的空間。如何能提升對不同語音文件類型的摘要的效能或品質，將是研究上一大挑戰。以下整理出數個語音文件摘要可能的未來研究方向：(1)以摘要的流暢性與可讀性而言，非摘錄式摘要仍舊表現的比摘錄式摘要優越，因此如何將目前的研究重心轉移至非摘錄式摘要，是一個重要的研究方向；(2)以摘錄式摘要而言，如何從語音文件中自動地訂定與切割摘要單位，將決定語音文件的摘要品質；(3)持續探索與發展更多不同類型的特徵以供監督式摘要模型使用；(4)由於監督式摘要模型的建立需有大量含人工標記的訓練語料集，故如何利用非監督式或半監督式(Semi-supervised)的方式來提升監督式摘要模型的效能、降低模型訓練時所需訓練語料集的時間，將是另一個研究重點[26]；(5)大部份以機器學習為基礎的摘要方法僅以考慮文件中個別每一語句是否屬於摘要語句的二元分類來做為模型訓練基礎，並沒有考慮到最後摘要結果之評估過程，因此如何發展以摘要結果評估之度量指標(Metric)為目標的摘要模型與訓練方法將是未來的可能研究方向之一[57]；(6)目前已被發展出的文件摘要效能的評估方式大多數僅是屬於詞彙層次上的比對，忽略了語意或主題層次上的比對。因此如何改良文件摘要效能的評估方式，使其能更貼切人類的評估方式亦將是一個具挑戰性的工作；(7)利用自動摘要結果輔助其它研究的進行，如資訊檢索、文件組織、或自動問答系統。

六、參考文獻

- [1] I. Mani and M. T. Maybury, *Advances in automatic text summarization*. 1999, Cambridge: MIT Press.
- [2] K. McKeown, J. Hirschberg, M. Galley, and S. Maskey, "From text to speech

- summarization,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2005, pp. 997 - 1000.
- [3] L. S. Lee and B. Chen, “Spoken document understanding and organization,” *IEEE Signal Processing Magazine*, **22**(5), 2005, pp. 42 - 60.
- [4] C. Chelba, T. J. Hazen, and M. Saraclar, “Retrieval and browsing of spoken content,” *IEEE Signal Processing Magazine*, **25**(3), 2008, pp. 39 - 49.
- [5] J. Carbonell and J. Goldstein, “The use of mmr, diversity-based reranking for reordering documents and producing summaries,” in *Proc. of Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1998, pp. 335 - 336.
- [6] J. J. Kuo and H. H. Chen, “Multi-document summary generation using informative and event words,” *ACM Transactions on Asian Language Information Processing*, **7**(1), 2008, pp. 3:1 - 3:23.
- [7] M. Galley, K. McKeown, J. Hirschberg, and E. Shriberg, “Identifying agreement and disagreement in conversational speech: Use of bayesian networks to model pragmatic dependencies,” in *Proc. of Annual Meeting of the Association for Computational Linguistics*, 2004, pp. 669 - 676.
- [8] C. D. Paice, “Constructing literature abstracts by computer: techniques and prospects,” *Information Processing and Management*, **26**(1), 1990, pp. 171 - 186.
- [9] M. Witbrock and V. Mittal, “Ultra summarization: a statistical approach to generating highly condensed non-extractive summaries,” in *Proc. of Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999, pp. 315 - 316.
- [10] H. P. Luhn, “The automatic creation of literature abstracts,” *IBM Journal of Research and Development*, **2**, 1958, pp. 157 - 165.
- [11] P. Baxendale, “Machine-made index for technical literature - an experiment,” *IBM Journal of Research and Development*, 1958, pp. 354 - 361.
- [12] H. P. Edmundson, “New methods in automatic extraction,” *Journal of the ACM*, **16**(2), 1968, pp. 264 - 285.
- [13] R. Barzilay and M. Elhadad, “Using lexical chains for text summarization,” in *Proc. of Workshop on Intelligent Scalable Text Summarization*, 1997, pp. 10 - 17.
- [14] T. Strzalkowski, J. Wand, and B. Wise, “A robust practical text summarization,” in *Proc. of AAAI Conference on Artificial Intelligence Spring Symposium on Intelligent Text Summarization*, 1998, pp. 26 - 33.
- [15] D. Marcu, *The theory and practice of discourse parsing and summarization*. 2000, Cambridge: MIT Press.
- [16] S. Teufel and M. Moens, “Summarizing scientific articles: Experiments with relevance and rhetorical status,” *Computational Linguistics*, **28**(4), 2002, pp. 409 - 445.
- [17] J. Zhang and P. Fung, “Extractive speech summarization using shallow rhetorical

- structure modeling,” *To appear in IEEE Transactions on Audio, Speech and Language Processing*.
- [18] Y. Gong and X. Liu, “Generic text summarization using relevance measure and latent semantic analysis,” in *Proc. of Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2001, pp. 19 - 25.
- [19] Y. T. Chen, B. Chen, and H. M. Wang, “A probabilistic generative framework for extractive broadcast news speech summarization,” *IEEE Transactions on Audio, Speech and Language Processing*, **17**(1), 2009, pp. 95 - 106.
- [20] R. Mihalcea and P. Tarau, “TextRank: bringing order into texts,” in *Proc. of Conference on Empirical Methods in Natural Language Processing*, 2005, pp. 404 - 411.
- [21] G. Erkan and D. R. Radev, “LexRank: graph-based lexical centrality as salience in text summarization,” *Journal of Artificial Intelligence Research*, **22**, 2004, pp. 457 - 479.
- [22] L. S. Lee, S. Y. Kong, Y. C. Pan, Y. S. Fu, and Y. T. Huang, “Multilayered summarization of spoken document archive by information extraction and semantic structuring,” in *Proc. of Annual Conference of the International Speech Communication Association*, 2006, pp. 1539 - 1542.
- [23] D. Wang, S. Zhu, T. Li, and Y. Gong, “Multi-document summarization using sentence-based topic models,” in *Proc. of Annual Meeting of the Association for Computational Linguistics*, 2009, pp. 297 - 300.
- [24] Y. L. Chang and J. T. Chien, “Latent Dirichlet learning for document summarization,” in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 1689 - 1692.
- [25] S. H. Lin and B. Chen, “Improved speech summarization with multiple-hypothesis representations and Kullback-Leibler divergence measures,” in *Proc. of Annual Conference of the International Speech Communication Association*, 2009, pp. 1847 - 1850.
- [26] S. H. Lin, B. Chen, and H. M. Wang, “A comparative study of probabilistic ranking models for Chinese spoken document summarization,” *ACM Transactions on Asian Language Information Processing*, **8**(1), 2009, pp. 3:1 - 3:23.
- [27] J. Kupiec, J. Pedersen, and F. Chen, “A trainable document summarizer,” in *Proc. of Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999, pp. 68 - 73.
- [28] G. Murray, S. Renals, and J. Carletta, “Extractive summarization of meeting recordings,” in *Proc. of Annual Conference of the International Speech Communication Association*, 2005, pp. 593 - 596.
- [29] J. M. Conroy and D. P. O’leary, “Text summarization via hidden Markov models,” in *Proc. of Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2001, pp. 406 - 407.

- [30] T. Joachims, *Learning to Classify Text using Support Vector Machines: Methods, Theory, and Algorithms*. 2002: Kluwer Academic.
- [31] M. Galley, "A skip-chain conditional random field for ranking meeting utterances by importance," in *Proc. of Conference on Empirical Methods in Natural Language Processing*, 2006, pp. 364 - 372.
- [32] D. Shen, J. T. Sun, H. Li, Q. Yang, and Z. Chen, "Document summarization using conditional random fields," in *Proc. of International Joint Conference on Artificial Intelligence*, 2007, pp. 2862 - 2867.
- [33] S. Maskey and J. Hirschberg, "Comparing lexical, acoustic/prosodic, discourse and structural features for speech summarization," in *Proc. of Annual Conference of the International Speech Communication Association*, 2005, pp. 621 - 624.
- [34] H. Christensen, Y. Gotoh, and S. Renals, "A cascaded broadcast news highlighter," *IEEE Transactions on Audio, Speech and Language Processing*, **16**(1), 2008, pp. 151 - 161.
- [35] K. Koumpis and S. Renals, "Transcription and summarization of voicemail speech," in *Proc. of International Conference on Spoken Language Processing*, 2000, pp. 688 - 891.
- [36] K. Zechner and A. Waibel, "Diasumm: flexible summarization of spontaneous dialogues in unrestricted domains," in *Proc. of International Conference on Computational Linguistics*, 2000, pp. 968 - 974.
- [37] Y. Liu and S. Xie, "Impact of automatic sentence segmentation on meeting summarization," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008, pp. 5009 - 5012.
- [38] J. Zhang, H. Y. Chan, and P. Fung, "A comparative study on speech summarization of broadcast news and lecture speech," in *Proc. of Annual Conference of the International Speech Communication Association*, 2007, pp. 2781 - 2784.
- [39] G. Penn and X. Zhu, "A critical reassessment of evaluation baselines for speech summarization," in *Proc. of Annual Meeting of the Association for Computational Linguistics*, 2008, pp. 470 - 478.
- [40] S. Furui, T. Kikuchi, Y. Shinnaka, and C. Hori, "Speech-to-text and speech-to-speech summarization of spontaneous speech," *IEEE Transactions on Speech and Audio Processing*, **12**(4), 2004, pp. 401 - 408.
- [41] Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper, "Enriching speech recognition with automatic detection of sentence boundaries and disfluencies," *IEEE Transactions on Audio, Speech and Language Processing*, **14**(5), 2006, pp. 1526 - 1540.
- [42] J. Hirschberg, "Communication and prosody: functional aspects of prosody," *Speech Communication*, **36**(2), 2002, pp. 31 - 43.
- [43] C. H. Wu, C. H. Hsieh, and C. L. Huang, "Speech sentence compression based on

- speech segment extraction and concatenation,” *IEEE Transactions on Multimedia*, **9**(2), 2007, pp. 434 - 437.
- [44] S. Maskey and J. Hirschberg, “Summarizing speech without text using hidden markov models,” in *Proc. of Human Language Technology Conference and the North American Chapter of the Association for Computational Linguistics Annual Meeting*, 2006, pp. 89 - 92.
- [45] S. Maskey, B. Zhou, and Y. Gao, “A phrase-level machine translation approach for disfluency detection using weighted finite state transducers,” in *Proc. of Annual Conference of the International Speech Communication Association*, 2006, pp. 749 - 752.
- [46] K. Knight and D. Marcu, “Statistics-based summarization - step one: sentence compression,” in *Proc. of National Conference of the American Association for Artificial Intelligence*, 2000, pp. 703 - 710.
- [47] M. Galley and K. McKeown, “Lexicalized Markov grammars for sentence compression,” in *Proc. of Human Language Technology Conference and the North American Chapter of the Association for Computational Linguistics Annual Meeting*, 2007, pp. 180 - 187.
- [48] F. Liu and Y. Liu, “From extractive to abstractive meeting summaries: can it be done by sentence compression,” in *Proc. of Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing 2009*.
- [49] *Text Analysis Conference (TAC)*. Available from: <http://www.nist.gov/tac/>.
- [50] C. Hori, T. Hirao, and H. Isozaki, “Evaluation measures considering sentence concatenation for automatic summarization by sentence or word extraction,” in *Proc. of Workshop on Text Summarization Branches Out*, 2004, pp. 82 - 88.
- [51] M. Hirohata, Y. Shinnaka, K. Iwano, and S. Furui, “Sentence-extractive automatic speech summarization and evaluation techniques,” *Speech Communication*, **48**(9), 2006, pp. 1151 - 1161.
- [52] H. Saggion and D. Radev, “Meta-evaluation of summaries in a cross-lingual environment using content-based metrics,” in *Proc. of International Conference on Computational Linguistics*, 2002, pp. 849 - 855.
- [53] C. Y. Lin. *ROUGE: Recall-oriented understudy for gisting evaluation*. 2003; Available from: <http://haydn.isi.edu/ROUGE/>.
- [54] F. Liu and Y. Liu, “Exploring correlation between ROUGE and human evaluation on meeting summaries,” To appear in *IEEE Transactions on Audio, Speech and Language Processing*.
- [55] E. Hovy, C. Lin, L. Zhou, and J. Fukumoto, “Automated summarization evaluation with basic elements,” in *Proc. of International Conference on Language Resources and Evaluation*, 2006.

- [56] A. Nenkova, R. Passonneau, and K. Mckeown, "The pyramid method: incorporating human content selection variation in summarization evaluation," *ACM Transactions on Speech and Language Processing*, 4(2), 2007, pp. 4:1 - 4:23.
- [57] S. H. Lin, Y. M. Chang, J. W. Liu, and B. Chen, "Leveraging evaluation metric-related training criteria for speech summarization," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2010.