5. 結論與未來展望

過去幾十年，有關自動文件摘要的研究主要仍著重於文字文件摘要；一直到1990年後期，由於影音多媒體技術的進步與成熟，才慢慢開始有語音文件摘要的研究。大部分的語音文件摘要方法多半由文字文件摘要方法延伸而來[Murray et al. 2005;Hirohata et al. 2005;Zhu and Penn 2005]；直至最近幾年，才紛紛有新的語音文件摘要方法開始被提出來[Kikuchi et al 2003;Furui et al. 2004;Wu et al. 2005;陳怡婷 et al. 2005;Chen et al. 2006;Maskey et al. 2006]。相較於一般傳統的文件摘要方法，本論文提出另一種模型架構來從事語音文件摘要，它同時亦適用於一般文字文件摘要。

文件摘要可分為摘錄式與非摘錄式摘要。本論文旨在探討摘錄式中文廣播新聞文件摘要方法。我們提出一個機率生成架構，它能將文句生成模型與文句事前機率緊密地耦合，用於摘錄式摘要之重要文句選取。將待摘要文件中每一文句被視為一個機率生成式模型，藉以預測文件生成的機率。我們提出二種機率生成模型：隱藏式馬可夫模型與關聯性模型的結合，以及詞層次混合模型，使用於文件摘要處理，並且經由一連串的實驗分析與討論，證明所提之方法的確可以較其他基礎實驗的摘要方法得到更高的正確率。同時，經由初步的實驗及實驗結果，可以看出所提之機率生成架構於語音文件摘要的運用，仍有很多進步及研究的空間，例如模型參數的設定、訓練。

此外，吾人也初步將文件結構特徵、語音辨識信心度與某些語音聲韻特徵使用於文句事前機率的估測。通常在一篇文件中，每一文句的重要程度都不相同；但是，其於文件中的重要程度資訊並沒有辦法直接取得。因此，嘗試以文句中某些摘要特徵的資訊來估測其事前機率值；我們於中文廣播新聞語料上進行一連串

的實驗，由初步的摘要結果證明某些摘要特徵，確實可以很好的估測出文句的事前機率分佈，同時提昇機率生成式摘要模型的摘要正確率。

基於這樣一個機率生成架構下，往後的研究將可對文句生成模型作進一步的改進，例如：1. 對於目前現有文句生成模型的改進，像是對文句機率分佈作更準確的估測，詞層次主題混合模型與關聯性模型的結果；2. 發展強健性之機率生成式模型參數的訓練與估測方式。亦或是提出其他的文句生成模型，以及其他文句事前機率的估測方法。其他研究方向，諸如進一步於機率生成式模型架構下，考慮摘要文句的重覆性及文句間關聯性的問題，如最大臨界相關摘要方法的概念；在選取重要文句時，除了考慮文句與文件之間內容的相關程度外，亦考慮文句與已摘錄之文句之間的相似度。