Lin-shan Lee, James Glass, Hung-yi Lee, Chun-an Chan, “Spoken Content Retrieval−Beyond Cascading Speech Recognition with Text Retrieval”, IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 23, No. 9, Sept. 2015, pp. 1389-1420.

This an overview paper in the journal. The overview papers were started in 2010 in this journal. It has to be on an important problem in the area, having co-authors from different groups to have a balanced view, can be much longer than a regular paper and at most 4 such papers each year. This paper gives an overview of the new concepts and directions of speech information retrieval, co-authored by a scientist in MIT.

Speech information retrieval refers to retrieving the multimedia directly based on the audio part of the multimedia information using speech instructions via hand-held devices, or the spoken version of Google. The approach used by most researchers in this area in the past was a cascading framework, i.e., performing speech recognition over the speech information and user instructions to transform them into text, followed by text information retrieval, as shown in Figure 1 of page 3, with performance highly dependent on recognition accuracy. Instead, this paper presents a whole series of new directions beyond the above cascading framework, e.g., learning the components of speech recognition and information retrieval jointly; utilizing the information lost in the recognition processes; or directly matching on the acoustic level rather than the word level to bypass the speech recognition. These approaches have performance less constrained by recognition accuracy, and have opened a wide space for further research.

這是該期刊的Overview Paper。該期刊自2010年才開始有這類論文，必須由跨團隊的作者群針對領域內之一重大問題合作完成以免偏頗，可以比一般論文長很多，每年至多只能有4篇；本篇論文由被提名人邀請MIT學者共同撰寫，主題是語音搜尋之新天地。

語音資訊搜尋是指根據多媒體資訊的語音旁白，使用語音指令藉助智慧型手機上網直接搜尋多媒體資訊(或語音版的Google)。此一領域過去多數研究者的作法，就是先將語音資訊及指令辨識成文字，後面串接文字資訊搜尋技術，但辨識錯誤就搜尋不對，如第3頁的圖1所示。本文指出了一系列新方向完全超越上述串接式的思維；例如將上述辨識及搜尋技術作整合性學習，或重拾語音訊號裡在辨識中漏失之資訊，或直接在聲音層面上比對不作辨識等；這些方向未必倚賴辨識正確，為此一領域開拓了遼闊的新空間。