行政院國家科學委員會專題研究計畫　期中進度報告

子計畫二：行動語音人機介面的研究與開發(2/3)

計畫類別： 整合型計畫
計畫編號： NSC94-2218-E-009-021-
執行期間： 94 年 08 月 01 日至 95 年 07 月 31 日
執行單位： 國立交通大學電信工程學系(所)

計畫主持人： 張文輝

報告類型： 精簡報告
報告附件： 出席國際會議研究心得報告及發表論文
處理方式： 本計畫可公開查詢

中　華　民　國 95 年 5 月 29 日

行政院國家科學委員會專題研究計畫報告

行動語音人機介面的研究與開發

ITS information access using voice over MANET

計畫編號：NSC 94-2218-E-009-021

執行期限：94 年 8 月 1 日至 95 年 7 月 31 日

主持人：張文輝　交通大學電信工程系 教授

一、中文摘要

(關鍵詞：語音對話系統，播放排程演算法。)

人性化的隨身資訊服務是智慧型運輸系統必備的功能，網際網路的興起更成為資訊傳播的重要平台，使用語音作為人機介面則可以提升行車安全與便利。本子計劃在MANET無線網路架構下，建構一行動語音對話系統，讓駕駛員以聲控操作取得道路指引及購物消費的生活資訊。本年度研究規劃針對MANET 的應用環境，設計製作一網路服務品質的量測平台，建立不同因素所對應的音質損害，再整合推導出一能正確反應通話品質的音質評量指標，以提供系統關鍵元件在錯誤控制與參數調整之用。在接收端，語音封包的播放緩衝器設計必須在播放延遲與封包漏失這兩者之間做一個權衡。針對這項需求，我們建立一個能具體反應網路電話的音質評量指標，以提供系統參數調整之準則。

英文摘要

(Keywords: spoken dialogue system, play-out buffer algorithm.)

The purpose of this three-year research is to develop a spoken dialogue system that allows drivers to use voice-controlled commands to access the ITS information server through a mobile ad-hoc network (MANET). This year, we focus on optimization of playout buffer in wireless speech transmission. Adaptive playout buffer algorithms rely on estimates of playout delay to compensate for variable network delays. In the proposed algorithm, the safety factor that controls the estimation process is dynamically adjusted according to a simplified version of the conversational-quality E-model. Perceptual based buffer design is formulated as an unconstrained optimization problem leading to a better balance between end-to-end delay and packet loss. Experimental results show that the proposed playout buffer algorithm can achieve better perceived speech quality than the basic adaptive algorithms.

二、計劃緣由與目的

行動語音人機介面採用雙向互動模式，遠端伺服器依辨認結果取得行車相關資訊，執行低位元率語音編碼處理，再經MANET 網路傳回到車內終端機作解碼播放。如同網路電話的工作原理，即時語音通訊必須架構在無連結式的UDP(User Datagram Protocol)傳輸協定，其缺點則是網路壅塞而封包漏失時不能要求重送。網路電話的服務品質取決於諸多因素，包括封包漏失、延遲時間、背景雜訊、及語音編碼失真，其中封包漏失率及延遲的容忍上限分別為10%及150msec。目前相關技術都採用錯誤控制碼(forward error control)及播放緩衝機制(playout buffer algorithm)。前者運用保護位元執行封包漏失的回復處理[1]，後者則彈性調整封包播放時間以對抗延遲擾動(delay jitter)[2]。這兩者之間其實存在因果循環關係，有效的錯誤控制碼所增加的延遲時間較長，進而影響播放緩衝時間的設定。除此之外，語音編碼處理有許多選擇[3]，如G.711 PCM、G.729 CS-ACELP、G.723.1 MPC-MLQ，而不同模式所衍生的信號失真及延遲時間亦存在明顯差異。較理想的系統設計是整體考量不同關鍵元件的最佳組合設計，且因應隨時變化的網路傳輸特性作合理調整。針對此項需求，首要之務為建立一個能具體反應網路電話音質的聽覺評量指標[4]，以提供系統關鍵元件在錯誤控制與參數調整之用[5]。問題是音質不容易得到一致而客觀的認定，產業界通常使用昂貴的檢測儀器來衡量。主要是因為音質評量牽涉到通道特性及系統架構兩層面，而不同服務品質因素所對應的音質損害程度更存在明顯差異。有鑑於此，我們將針對Voice over

MANET 的特定應用環境，設計並製作一正確量測各項服務品質因素的測試平台[6]。同時參考國際電信聯盟ITU 針對網路系統規劃所制定的E-model[7]，根據聽覺評量建立不同因素所對應的音質損害，再整合推導出單一能具體反應網路通話品質的音質評量指標。

三、研究方法與結果

計畫主要設計製作一網路服務品質的量測平台，以提供系統關鍵元件在錯誤控制與參數調整之用。行動語音對話系統採雙向互動的模式，遠端伺服器經由無線網路接收語音特徵參數，再依辨認結果回傳相關資訊給行動終端機作解碼播放。因應時變的無線網路服務品質，我們採用錯誤控制碼作封包漏失的回復處理，同時彈性調整封包播放時間以對抗其延遲擾動，系統流程如圖一所示。系統整合成敗的關鍵在於明確掌握MANET 傳輸特性且建立其對應的數學模型，方能因應不同車速及應用環境作合理規劃。

(1) 音質評量指標的制定

不同的網路服務品質參數所對應的音質損害，會因為通道特性及系統架構兩層面而存在明顯的差異，傳統的音質評量難以達到一致且客觀的標準認定。因此我們參考國際電信聯盟ITU 針對網路系統規劃所制定的E-model，配合量測所得的封包漏失率及延遲時間來制定音質評量指標。推導過程描述如下：首先利用量測結果比對得到延遲時間與音質損害參數的對應關係，再建立不同語音編碼處理所對應的音質損害值。這是因為語音編碼處理依位元率區隔而有許多選擇，如G.711 PCM、G.729 CS-ACELP、G.723.1 MPC-MLQ，不同模式所衍生的信號失真亦存在明顯差異。除此之外，網路傳輸常常會因為網路擁擠或其他不可預知的因素導致封包漏失，而不同的漏失率所對應的音質損害也必須納入設計考量。藉由完成不同音質損害的估測，即可整合推導出一項能具體反應網路通話品質的音質評量指標，進而提供系統關鍵元件在錯誤控制與播放控制參數調整之用。

(2) 封包漏失與延遲的估測

語音封包傳輸過程經由Sniffer 軟體分別在傳送端與接收端蒐集到每個封包之識別碼以及抵達時間差等資訊，將用於估測封包漏失與延遲擾動(delay jitter)等網路服務品質參數。由於延遲擾動會受到傳送與接收兩端電腦時脈誤差的影響，而無法獲得準確的估測，因此我們提出一有效去除時脈誤差的估測演算法，以期每個封包量測的準確度能維持在50 微秒誤差範圍之內。其演算過程描述如下：首先利用序列識別碼來判定封包的漏失並將接收封包依序排列，配合傳送與接收序列所對應的抵達時間差計算其累加相差時間，再利用此時間與累加封包數目的近似線性關係進行線性逼近，所求得的斜率即可用於去除兩端電腦的時脈誤差，進而提昇延遲擾動的估測準確度，也用以提升品質量測與播放緩衝演算法的精確度。

(3) 播放緩衝機制

語音信號會以固定的間隔來產生封包，透過網路傳送到接收端。每個封包的網路延遲會取決於所走的路徑頻寬以及該路徑上路由器的壅塞程度，而網路延遲的差異即為延遲顫動。為了降低顫動的影響，接收封包在播放前會先被暫存在一緩衝器一小段時間，藉此減少封包漏失的機會。播放延遲被設計為 $d_{play,i} = \hat{d}_i + \beta \hat{v}_i$，其中 $\hat{d}_i$ 與 $\hat{v}_i$ 由過去所紀錄的封包網路延遲資訊來估算，而 β 是一個安全緩衝因子。較大的 β 值會導致一個較低的封包漏失率，然而播放延遲會因此增加。現存演算法中的 β 值是固定大小，並未考量通話音質。為了使播放緩衝機制能夠因應時變的網路特性，需建立網路延遲與封包漏失所對應的音質損害模型，有效的整合於單一具體反應通話音質的評量指標，進而提出音質最佳化的適應性播放排程機制。音質評量模型

$$\text{(E-model)}: R = 94.2 - I_d(d) - I_{ec}(r) - I_{epl}(e)$$

利用統計方法分析網路延遲並建立封包漏失與播放延遲的關聯性，將音質最佳化的適應性播放排程機制轉換為一個最佳化的

問題，針對每個封包的播放延遲適應性的調整 $\beta_i^* = \arg\max_{\beta}\{R\}$。

(4) 系統結構的最佳化設計

播放排程演算法為了達到較低的漏失率常造成較長的播放延遲，而漏失率與平均播放延遲存在著一個取捨(trade-off)的關係。考慮因應網路環境的變化進行播放排程調整，較合理的方法彈性地修改播放排程演算法的安全因子 β。目的是希望能夠因應時變的網路特性，適度的去權衡封包漏失率以及平均播放延遲之間的關係進而動態地調整安全因子 β 值。而進一步考慮音質效果，搭配彈性的播放排程以封包為單位來調整播放時間，運用音長調整機制延伸或壓縮播放時間，用以補償可適性播放緩衝機制中過度延遲的封包。音長調整機制如圖二所示，利用正弦分析模型解析聲帶激發訊號以及發聲共振腔系統模型，利用聲帶訊號延長和壓縮激發聲音訊號並考慮合成訊號在不同封包所造成聲音不連續的現象，藉由相位連續性的估算，產生連續清晰的聲音訊號。因此利用音長調整將可以精確的配合時變性播放緩衝的機制。

四、實驗結果與討論

透過正確的量測網路延遲與漏失，運用音質評量指標評估語音傳輸系統的優劣程度。針對網路傳輸影響的程度，我們考慮使用彈性的錯誤控制碼、可適性的播放緩衝估測演算法以及時變的音長調整，可充分補償過度延遲或漏失封包所造成嚴重的音質衰減問題。在實驗中針對 AR-based 與 NLMS-based 這兩種播放排程演算法的效能做比較，如表一所示。其中 β 值分別設定為 4，6 與動態調整的 β_i。由表一可發現，AR-based 演算法中不存在一個固定的 β 值，在面對不同的網路環境皆可達到最好的音質表現。例如; β = 6 適用於 light 的網路環境，β = 4 適用於 heavy 的網路環境。模擬的結果顯示，適應性的調整 β_i 可強化現存的演算法。在 heavy 的網路環境，一個動態的 β_i 值得音質評量分數比 β =4 或 6 來的音質評量分數高。主要是因為適應性的調整 β_i 值可以達到延遲與封包漏失間的最佳權衡值。

五、結論

藉由分析網路通道所造成的漏失及延遲，進行音質評量的鑑別，可提供使用者對於網路環境優劣程度以及傳輸聲音品質有一客觀的衡量準則。進一步利用延遲與漏失的量測，設計音質最佳化的語音封包播放緩衝機制。由實驗結果得知可適性的播放緩衝機制配合音長調整機制，可以在延遲與漏失的權衡下，在聽覺品質的量測中獲得較好的效果。

六、具體成果

此研究結合隨意行動網路與分散式語音辨認平台，開發一車用旅行資訊檢索系統。本年度針對語音在無線網路所面臨的傳輸狀況量測、音質評量指標的建立以及播放緩衝機制等問題進行探討。具體研究成果將在下列國際語音通訊相關研討會發表。

[1] C. F. Wu, I-Te Lin, and W. W. Chang, “Adaptive playout scheduling for multi-stream voice over IP networks” accepted to present in 2006 European Signal Processing Conference.

[2] C. F. Wu, and W. W. Chang, “Perceptual optimization of playout buffer in VoIP applications” submitted to Internet Conference on Communication and Networking in China.

[3] I-Te Lin, C. F. Wu, S. H. Chen , and W. W. Chang, “Multiple description quantization for recognizing voice over packet Networks” submitted to Internet Conference on Communication and Networking in China.

[4] C. L. Lee, and W. W. Chang, “Symbol-Based Source-Controlled Channel decoding and its Application to Distributed Speech Recognition,” submitted to International Conference

on Information Sciences.

七、參考文獻

[1] J. C. Bolot, S. F. Parisis, and D. Towsley, “Adaptive FEC-based error control for Internet Telephony,” in *Inforcom’99*, March 1999.

[2] S. B. Moon, J. Kurose, and D. Towsley, “Packet Audio Playout Delay Adjustment: Performance Bounds and Algorithms,” ACM/*Springer Multimedia Systems*, vol. 5, pp. 17-28, January 1998.

[3] L. Hanzo, F.C. A. Somervilie, and J.P. Woodard, Voice Compression and Communications. Wiley-Interscience, 2001.

[4] M.E. Perkins, C.A. Dvorak, B.H. Lerich, and J.A. Zebarth, “Speech Transmission Performance Planning in Hybrid IP/SCN Networks, “*Communications Magazine, IEEE* , Volume: 37 , Issue: 7 , July 1999, pp. 126 – 131.

[5] W. Jiang and H. Schulzrinne, “Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss,” in *Proc. ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video*, May 2002, pp. 73-81.

[6] J. Feigin and K. Pahlavan**, “**Measurement of characteristics of voice over IP in a wireless LAN environment,” in *Proc. IEEE International Workshop on Mobile Multimedia Communications*, Nov. 1999, pp. 236 – 240.

[7] “The E-Model, a Computational Model for Use in Transmission Planning, ” ITU-T Recommendation G.107May 2002.

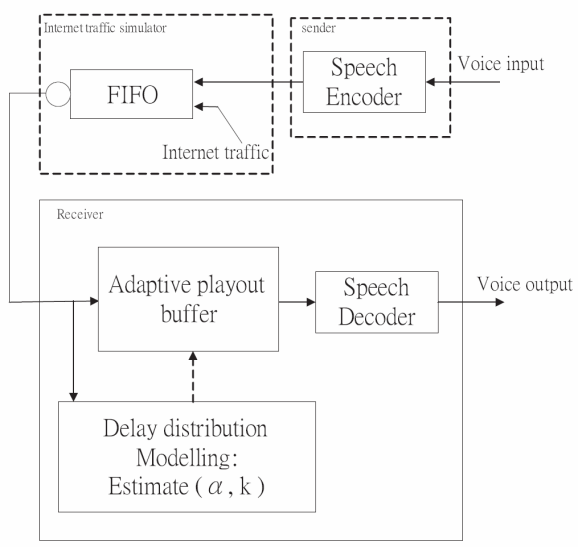

圖 1：語音傳輸系統方塊圖。

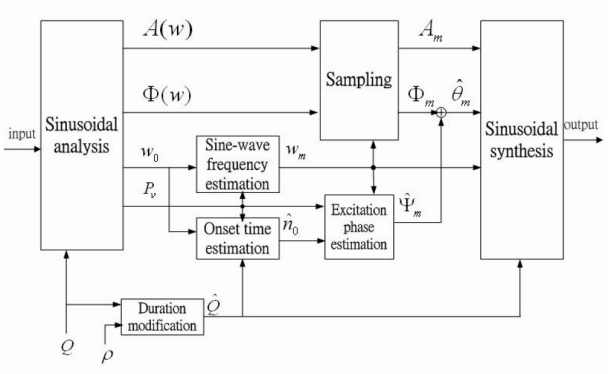

圖 2：音長調整機制方塊圖。

表 1：不同播放緩衝演算法之結果。

Trace	Playout algorithms	Loss %	Delay (ms)	MOSc
Heavy	AR, $\beta = 4$	3.91	281.85	2.666
	AR, $\beta = 6$	3.40	299.60	2.615
	AR, dynamic β	2.84	296.93	2.728
	NLMS, $\beta = 4$	6.79	272.46	2.339
	NLMS, $\beta = 6$	3.61	289.88	2.675
	NLMS, dynamic β	2.56	298.31	2.765
Light	AR, $\beta = 4$	2.42	70.45	3.723
	AR, $\beta = 6$	2.09	79.04	3.760
	AR, dynamic β	2.06	73.03	3.776
	NLMS, $\beta = 4$	2.33	65.35	3.746
	NLMS, $\beta = 6$	2.11	68.12	3.774
	NLMS, dynamic β	2.04	73.65	3.778