大型網路語音會談中回音消除方法

Echo Cancellation In Large-Scale VoIP Conferencing

連耀南
國立政治大學 資訊科學系
lien@cs.nccu.edu.tw

祁立誠
國立政治大學 資訊科學系
g9613@cs.nccu.edu.tw

摘要

在多人參與網路會談時，因為聲音在空間中傳遞或反射等因素，使得由喇叭發出的聲音再次被麥克風收回，造成回音的產生。只要有一位使用者的裝置產生回音時，回音訊號就會在與會者之間擴散，使得所有使用者受到影響。此狀況在與會人數越多時，發生機率越高，且影響越嚴重。

由於網路會談沒有標準的聽筒設備，使得回音延遲的時間難以預估，且因為網路傳輸與聲音失真等因素，導致傳統的回音消除機制在多人網路回談中經常失效。

本研究提出藉由語音動態偵測(Voice Activity Detection-VAD)的方式分辨回音訊號，藉由本研究所提出的語音能量 VAD 判定機制，能有效區別正常語音與回音的差異，即可有效的消除回音，同時發揮靜音抑制(Slience Suppression) 的效果，阻擋不含語音內容的封包，降低網路頻寬耗用。本研究以自行開發的 VoIP 軟體進行實地測試實驗，實驗中顯示，我們的方法能消除 85%以上的回音。

一、簡介

近年來，為響應全球減碳運動，並節省差旅費，許多公司內部召開之大規模跨國會談均採用網路會談(conference)方式進行。當參與會談人數增加時，許多網路會談的問題就顯得更為嚴重，回音(echo epidemic)即為其中一項嚴重的問題。

1.1 回音現象

聲音由音源發出後，若被反射回音源處即產生回音。若第三者直接接收到音源發出之聲音後，又接收到反射的回音，則接收到兩次以上的相同聲音。回音的情形在全雙工(full-duplex)語音通訊系統中經常發生。由於全雙工系統能夠同時傳送與接收聲音訊號，因此收聽者(listener)本身也同時為說話者(talker)。而聲音可在固體與空氣中傳導，如此可能使收聽者的收音裝置收到自己的播放裝置所發出的聲音，回音的現象就此產生。如圖 1 即為傳統電話中的回音產生情形。

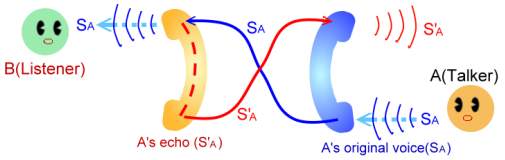

圖 1：基本回音現象

1.2 回音消除基本原理

若要消除前述的由聲音上傳輸造成的回音，最簡單的方式就是加入回音消除機制(echo cancellation)。對傳統電話而言，由於有標準的聽筒裝置，因此很容易預估回音的強度與傳遞所需要的時間，建構此回音消除機制並不困難。

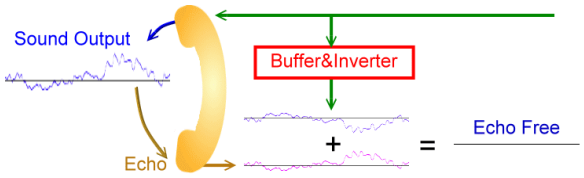

圖 2：傳統電話中的基本回音消除機制

如圖 2，聲音在由聽筒播放之前，先存入暫存器之中，並且加以反向(inverter)，當話筒收回聲音訊號之後，再將先前存下的訊號經過適當衰減後與此相加，即可消除由話筒收回的回音[1]。

1.3 VoIP 中的回音

在 VoIP 通話進行時，很多使用者並不會特別配備專用的設備，而是採用電腦原有的喇叭與麥克風進行網路通話。如此一來聲音由喇叭播放後，在空氣中傳輸，經過空間反射後再次被麥克風收回，同樣會產生回音，此種現象的發生將可能影響通話的品質，嚴重時甚至影響會談的進行。例如圖 3 中說話者(A)的聲音傳送至收聽者(B)的喇叭播放後，經過牆壁反射，又被麥克風收回，返回至 A 的喇叭播放，造成回音。聲音自喇叭至麥克風的路徑稱為”Accoustic path” 。

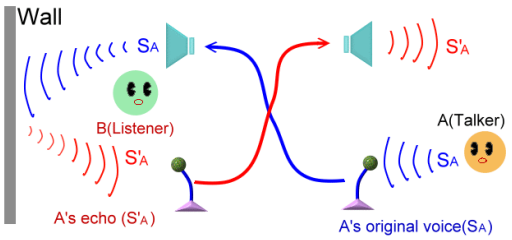

圖 3：因為聲音的傳輸或反射而造成的回音

在多人網路會談時，若其中一位與會者的喇叭與麥克風距離太近，使得由喇叭放出的訊號被麥克風收回。當回音消除機制正常啟動時，此訊號在裝置可被順利消除，而不至於對網路會談造成任何影響。然而電腦軟硬體都有失效的機會，使得消除機制未必能夠正常運作，多人會談中只要有一個參與者的回音消除機制失效，成為回音的產生者而產生回音時，則可能對整個會談造成影響。如圖 4，在會談中，有其中一部裝置(B)的喇叭發出的聲音被麥克風收音，且該裝置的回音消除機制並未正常工作，這將使得 B 成為回產生者(即 Echo Generator)。

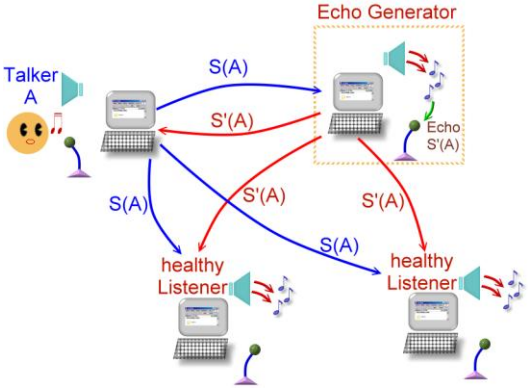

圖 4：一個 Echo Generator 的回音影響

只要有任何一個裝置的回音消除機制故障，則其產生的回音就會干擾發話者以及其他所有收聽者，造成整個網路會談充滿回音的聲音訊號。更嚴重的情形下，在一個會談中有超過一個以上的 Echo Generator 存在時，將不斷重複產生回音，直到聲音衰減至無法收音為止，此時只要有一位使者說一句話，所有與會者就會不斷聽到重複遞迴的回音訊號，使會談難以進行。

1.4 Proximity Problem 造成的回音

所謂的 Proximity Problem 為電腦之間距離太近時，所發生一種聲音傳遞上造成的干擾問題[5]。若兩部電腦距離太接近時，則使用者的聲音（假設只有一個人說話）及兩部電腦的喇叭聲音，同時被兩部電腦的麥克風收回，此狀況發生的狀況如圖 5 所示。兩部電腦的喇叭將會同時放出包含發話者以及不斷重複的回音，且如此混亂的聲音將再次被雙方的麥克風收音，導致此現象不斷循環。如此一來，將會造成複雜且混亂的回音加上互相收音所造成的雜音。就如雪崩效應一般，收回的聲音被放大後播出，接著又立刻被收音且播放，不斷反覆而造成類似震盪的刺耳噪音。

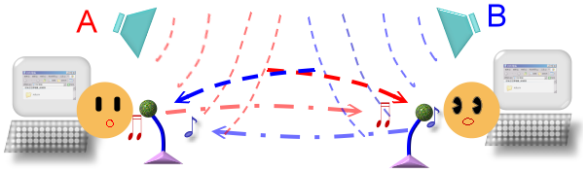

圖 5：Proximity Problem 現象

目前各種針對回音消除的方法，大多僅針對一對一的通話所產生的回音進行消除，且考慮的僅有聽筒到話筒間固定長度且短距離造成的回音，針對距離不固定，與會人數增加時所造成的回音狀況常常無法正確的消除。此外，當與會成員沒有在說話時，傳統的會談程式仍然會持續將空白的語音封包送至網路上，造成網路資源的浪費。本研究考慮的問題為如何改善 echo epidemic 對於通話品質所造成的影響，同時作到靜音抑制(Silience Suppress)降低總頻寬使用量。本研究分析大型網路語音會談中，回音消除機制失效的成因與語音的特性，以此設計消除回音與靜音抑制的方法，抑制會談中的回音與靜音，確保會談正常進行

二、背景與相關研究

1950 年代以前，電話系統尚未有多方通話技術出現，訊號傳輸以半雙工(half-duplex)方式進行，同一時間僅其中一方在說話，另一方是收聽者[11]。當時沒有回音消除機制存在，為了消除回音對通訊造成的干擾，僅採用回音抑制(echo suppression)的方式降低回音。此機制會判定電話的那一方是說話者，則在說話期間就保留正常的語音，將收聽者回傳的訊號視為回音進行衰減或阻止其傳輸，以達到回音抑制的目的。直到 1970 年代，隨著半導體的進步，市場上才逐漸出現回音消除機制的產品。此時的技術開始採用訊號暫存與相減的方式消除回音[7]。

2.1 回音消除原理

在 VoIP 系統中，不像傳統電話，由於沒有標準的聽筒裝置，因此回音的延遲時間與音量失真狀況相當難以預估。所以針對此狀況需要加入更多的判斷機制來掌握回音的狀況，並即時的進行消除。如[2]文中所提出的解決方案，此系統在受話端加入回音消除裝置，且包含了訊號即時比對的能力：

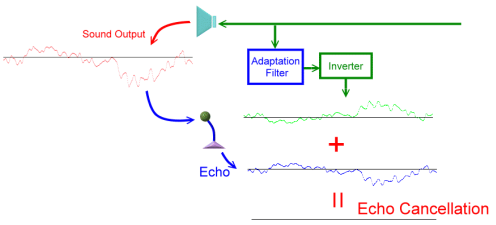

圖 6：VoIP 系統中回音消除機制

此系統透過 LMS/NLMS 等演算法[6]，將可能有回音的訊號做誤差還原後，逐一與原始訊號做比較，判斷回音是否發生。若回音發生時，則將事先暫存的原始訊號取出，經過處理後相減以消除回音。但此裝置需要相當可觀的運算資源才能夠完成，因此通常以 DSP 數位訊號處理器實做，成本居高不下。

2.2 回音消除方法分類

回音消除機制根據實做的使用者端不同，可分為以下兩種方式：

2.2.1 Listener Echo Cancellation

這是『由回音產生者消除回音』的方法，亦即由麥克風擷取聲音(發話者)一方自己進行回音消除的動作。若此方法使用在一對一的傳輸狀況下則稱為 Near End Echo Cancellation。

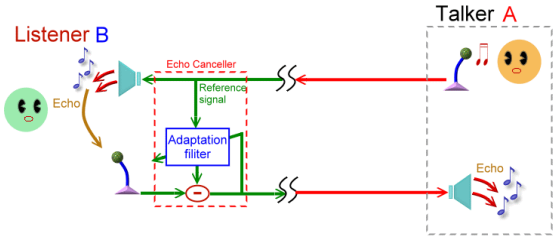

圖 7：Listener Echo Cancellation 機制

如圖 7，此方法是由產生回音的收聽者 B 負責消除回音。當收聽收到來自發話者 A 的訊號時，會先經過暫存。當麥克風收到回音時，就將此包含回音的訊號與原本暫存反向後的訊號相減，以去除回音。在 VoIP 系統中，雖然每部參與會談的裝置的都有 Listener Echo Cancellection，但卻常因某些因素造成此機制失效，這些原因包含：

- VoIP 沒有標準的話筒與聽筒，回音延遲時間與音量難以估計。
- 回音可能經過牆壁反射而造成相位相反，此時若再與反向過的訊號相加，反而使回音更為嚴重。
- 由於環境(空間，材質，距離等)因素，麥克風收到的回音訊號可能會在相位與頻率上產生嚴重失真。

2.2.2 Talker Echo Cancellation

與前述相反，這是『由回音接收者消除回音』的方法。由聽到回音的收聽端在播放聲音之前，先將聲音中的回音去除再播放。若此方法使用在一對一的傳輸狀況下則稱為 Far End Echo Cancellation。[3][4]

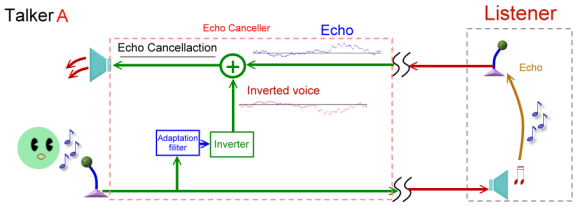

圖 8：Talker Echo Cancellation 機制

如圖 8，此方法是由說話者(A)在送出聲音之前，先將自己的聲音存下，當回音訊號透過收聽端傳回後，由說話者 A 取出先前暫存的訊號，比對回音是否存在，一旦發現回音存在，就將訊號經過衰減後相減。

若要設計一套由 Talkerer 端主動消除回音的機制有相當多挑戰存在，尤其在多人網路會談中，某些問題將變得比一對一時更加複雜，這些問題列舉如下：

- 使用者之間距離相當遙遠(例如跨國 VoIP 會談)，則回音到達的時間可能難以預估。
- 同時多人講話時，多個人聲混合後特徵值不明顯，使聲音特徵比對難以進行，判斷回音是否存在。
- 封包傳輸時，有可能因為網路因素，造成封包不依照順序到達(即 time sequence disorder 的問題)，導致回音和正常語音出現時間混亂，使回音判定演算法失效。
- 運算時間之限制與混音器造成的混淆。
- 多人會談時，使用者們有些說話行為會造成誤判。

雖然目前的 VoIP 軟體與電腦音效裝置均提供回音消除機制，但卻經常失效(或回音消除不完全)而導致回音產生。本研究的目標即為在不增加額外硬體設備與運算負擔的前提下，提出一套回音消除機制的方法，並可兼具靜音消除的功能，大幅降低網路會談的頻寬需求。

三、MET VAD 靜音及回音消除機制

本研究提出的解決方法為在每一部參與語音會談裝置加入一套判定麥克風收到的聲音是否為正常說話語音的機制。當沒有回音產生或回音消除機制正常運作時，就將聲音以正常方式送至網路中。反之，若有回音存在或使用者沒有說話，導致麥克風聲到回音或是靜音訊號時，就由此機制阻擋訊框。

3.1 VAD 語音動態偵測

語音動態偵測 VAD (Voice Activity Detection) 的目的為找出聲音訊號中，實際含有語音內容的區段。

一個有效的 VAD 方法是利用語音的頻率特性做判斷，但因回音之頻域特性與正常語音類似，故不適用於回音消除。另一法為根據能量大小作為判斷依據。其技術上的作法為：定義每一個聲音訊框的能量值(energy)，同時設定一個臨界值(threshold)作為判定用的依據，其判定的演算法如下(其中，Ej 為能量值，Er 為臨界值)：

IF (Ej> Er)
THEN Frame is ACTIVE
ELSE Frame is INACTIVE

當一個聲音能量的能量值超過臨界值時，即將此訊框視為語音。反之，即視為非語音。而本研究將以能量作為判斷為正常語音或回音的依據。

3.2 系統架構

本研究提出以 VAD 的方式作為正常語音與否的判定機制。若使用者的終端設備原內建有回音消除機制，此架構也能與現有之 Listener 回音消除機制搭配共同運作，當原有回音消除機制失效時，VAD 仍然能有效阻擋回音與靜音訊框。此架構如圖 9 所示。

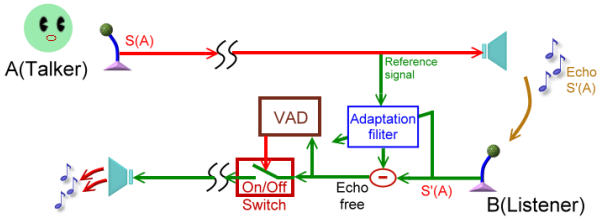

圖 9：加入 VAD 機制之系統

若回音消除機制失效，造成收聽者 B 的回音無法正常被消除時，此 VAD 機制即發揮功能：

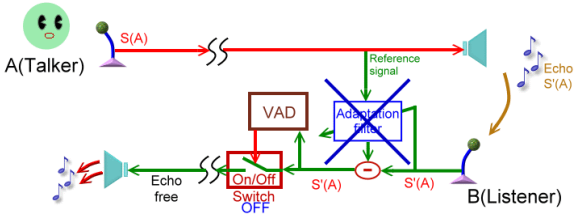

圖 10：加入 VAD 機制，且回音消除機制失效時的狀況

而當 B 端的 Listener Echo Cancellation 失效，導致回音之產生，但同時 B 也在說話，亦即 B 的正常語音 S(B)與說話端 A 的回音 S(A)夾雜在一起被送出時，VAD 機制能判斷出語音訊號存在，並不會將此聲音擋下，仍然會送出此聲音訊號，而使溝通能正常進行(但回音並未被消除)，此狀況如圖 11：

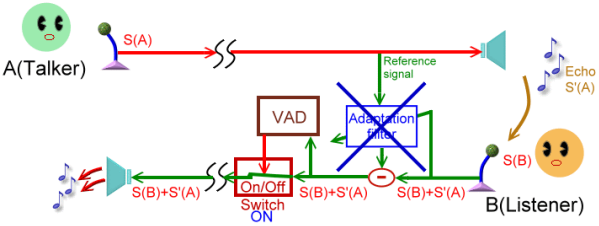

圖 11：回音消除機制失效，同時收入回音與說話聲音時的狀況

3.3 演算法設計

傳統 VAD 為了判斷人聲與否，常用時域(time domain)與頻域(frequency domain)兩種特性判斷方法。由於人類的聲音必然集中在特定的頻率範圍，因此採用頻域判定通常能得到比較好的判定結果。但為了分析頻域數值，輸入的每一個訊框都必須先轉換為頻域數值，此部份的運算量相當可觀。在本研究中，根據實驗可以發現能夠精確判斷人聲與否的頻域 VAD 方法並不適用於回音判斷，其原因為無論語音或回音，聲音的頻率範圍均相同(來源都是說話聲音)，在此狀況下，採用運算複雜度較低的時域判斷反而能夠有較好的判定結果。

在使用者進行網路會談的實際環境中，由於喇叭音量會被使用者調至適當大小，由喇叭發出再次被麥克風收回的回音通常會經過空間上的傳遞與反射，造成能量的衰減，再次進入麥克風時，其音量比正常使用者說話的聲音小。

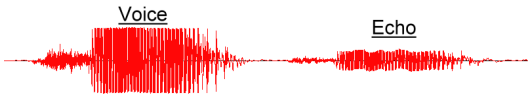

圖 12：正常語音與回音的振幅差異

圖 12 即為在一段聲音訊號中，正常語音與回音的音量比較。一般的會談中，使用者必然面對麥克風直接說話，因而麥克風得到的音量(即振幅)必然較大，相反地，回音由於經過傳遞時的衰減，因此振幅較小。由於具有此種信號特性上的差異，即可做為時域能量 VAD (Time Domain Energy-Based VAD) 的判定依據。

在時域(Time-Domain)訊號分析中對於聲音訊框的能量值(Energy)定義如下式[9]：

$$E = \sum_{k=0}^{N} S^2[k] \cdots\cdots (1)$$

其中：

E 為訊框的能量值。
N 為訊框的取樣總數。
S [k]則為第 k 個取樣的振幅。

亦即將一個聲音訊框中每一個取樣的振幅平方值加總後，即為該訊框的能量值。而本研究的設計中，每當由麥克風收到的聲音經過取樣後，都先經過上式計算出該訊框的能量值，以作為後續 VAD 演算法的判定依據。

3.3.1 LED VAD 演算法

LED VAD 即為『線性能量偵測』(Linear Energy-Based Detector)，是一種最常用來定義並更新能量判斷臨界值的方法[8][10]。其演算法的初始臨界值(Initial Threshold)設定為第一個訊框的能量值：由於假設第一個訊框的內容必然為非語音，故此訊框的能量值作為背景雜訊的初始值。接下來收到每個訊框號，對臨界值更新如下式：

$$E_{dnew} = (1-p)\cdot E_{dold} + p\cdot E \cdots\cdots(2)$$

其中：
Ednew 為每次更新後的臨界值。
Edold 為前一次的臨界值。
E 為最近一次的訊框能量。

而上式中的 p 則為可調整的參數，可根據不同聲音環境或需求做調整。當 p 越大，則臨界值更新越快速。但由於 LED VAD 演算法的臨界值變化隨時跟著聲音能量(即振幅大小)變動，因此除非使用者非常頻繁的說話，否則當正常語音中斷時，回音的訊號極容易被誤判為正常語音訊號，降低回音消除之效率，因此並不適用於回音偵測。

3.3.2 MET VAD 演算法

為了改善前述 LED VAD 演算法所遇到的問題，必須針對語音會談的回音判定演算法重新進行設計。本研究針對多人參與的網路語音會談進行實驗與分析，大多數情況下會有一些共同的特點：

- 回音的能量值大部分都在正常語音能量的 10%~15%以下。
- 與會人數越多時，平均每位使用者發言的機會就越少。
- 每一位使用者會在上線後的短時間內發言，以告知其他使用者(例如打招呼)，因此會談程式在啟動後短時間內應該會收到一段正常語音訊號。

根據以上的特點，本研究設計了一套專門針對回音偵測的 VAD 判定演算法，稱為『最大能量追蹤 VAD』(Maximum Energy Tracking VAD - MET VAD)。此方法可以確保使用者於持續一段時間未說話時，仍然不會將回音誤判為正常語音。MET VAD 演算法之實際步驟如下：

a. 開始進行判斷之前，可由程式根據先前的設定值，作為初始臨界值。若無先前之臨界值，則以第一個訊框能量值的 10 倍作為初始臨界(假設第一個訊框內容必然為非語音)。

b. 每個訊框輸入後，判斷其能量是否大於臨界，若是則視為語音，否則為非語音。

c. 若輸入訊框能量之 15%超過目前之臨界值，則更新臨界值，以此訊框能量之 10%作為新的臨界值。

d. 為了避免因為雜訊導致臨界值設定過高，當超過 100 個訊框沒有收到語音訊框時，則將臨界值降低為 95%。

根據以上原則，此判定機制會抓取輸入的最高能量值，並取此值的 10%作為臨界值，由於絕大多數回音訊號都低於此能量，故可以依據每一位使用者的說話能量，過濾掉回音。另外為了避免因雜訊或其他因素導致臨界值設定過高，故加入自動降低臨界值的機制：當長時間沒有收到正常語音時，就略微調低臨界值，避免使用者因為意外而無法送出聲音。圖 13 為 MET VAD 演算法之判定流程圖，其中 t 為臨界值決定參數，在此定為 10%。

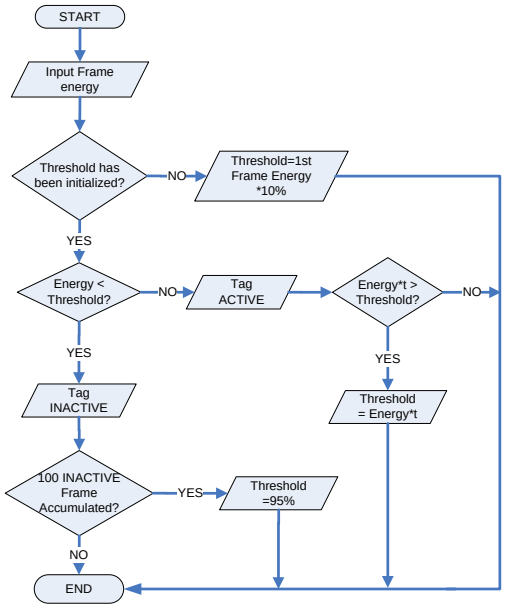

圖 13：MET VAD 演算法流程圖

對此演算法而言，整體的時間複雜度僅為線性時間 O(n)。以取樣率 8KHz 來說，每秒 8000 次取樣的運算量負擔相當輕微。此流程之虛擬碼 (pseudo code)如下：

```
NewFrameArrival(){
IF(Threshold not been Initialzed)
    Threshold = 1st FrameEnergy*10%;
 ELSE IF(FrameEnergy > Threshold){
     TagACTIVE();
     IF(FrameEnergy*t > Threshold)
         Threshold=FrameEnergy*t;
}
ELSE IF(FrameEnergy < Threshold)
    TagINACTIVE();
IF(100 INACTIVE Frame Accumulated)
      Threshold=Threshold*0.95
}
```

四、效能分析

我們針對本研究所提出的方法進行實際驗證，首先以樣本資料對 MET VAD 與現有的其他 VAD 進行量化評比，其次，於真實的網路會談程式中進行質化驗證。

4.1 評估指標

本研究使用誤判率及 MOS 作為評量指標。

4.1.1 誤判率

誤判率可分為兩種：False Positive (把回音當成正常語音的誤判)與 False Negative (將正常語音當成回音的誤判)。若 False Positive 高，則通話聲音中剩餘的回音將干擾會談，若 False Negative 高，則正常的語音會被 VAD 刪除，使用者溝通可能受阻。

4.1.2 MOS

由於聲音品質對於使用者而言是因個人主觀感受而異的。每個人在進行語音會談時，對於語音失真與回音程度的可接受範圍均不相同。實際將聲音資料經過 VAD 處理後，由使用者聆聽並評估其對回音的過濾能力以及對語音資訊的破壞程度，作為評估 VAD 效能的依據。MOS 大於 3 時代表能夠正常溝通，而小於 2 時代表溝通困難。

4.2 實驗一：以聲音樣本評比各種 VAD

本實驗事先製作了一段聲音樣本資料作為測試樣本，其中包含大部分的回音與部份的語音，並且事先分析語音的位置，長度與能量等資訊。本實驗採用的樣本聲音資料如表 1：

表 1：實驗一使用的樣本聲音資料

聲音取樣率	8000 Hz
取樣格式	Mono PCM
取樣位元數	8 bits (256 Levels)
聲音長度	15 秒
訊框長度	30ms
總訊框量	500
語音插入位置	1.5 秒 (第 50 個訊框)
語音插入長度	1.2 秒 (共 40 訊框)
回音能量峰值：語音能量峰值	22%

圖 14 為本實驗輸入聲音之波形：

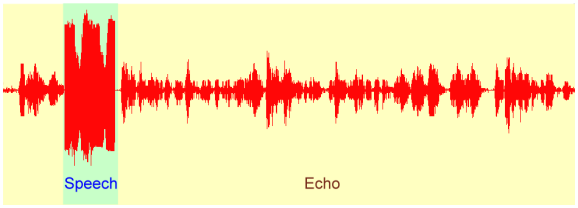

圖 14：實驗一輸入之聲音波形

本實驗使用 LED VAD，WFD VAD 與 MET VAD 三種時域能量 VAD 演算法判定以上的輸入樣本訊號。

4.2.1 LED VAD 演算法

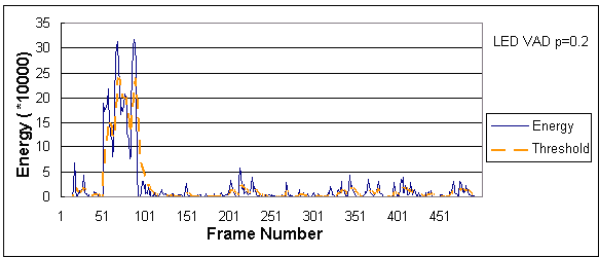

圖 15：以 LED VAD 演算法針對樣本資料的分析結果(p=0.2)

表 2：LED VAD 演算法的誤判率

參數 p	p=0.1	p=0.2	p=0.3
False Positive (%)	31.3	33.91	33.48
False Negative(%)	30	45	50

實驗結果可看出 LED VAD 演算法的臨界值會隨著輸入能量的大小而變化臨界值，因此當聲音能量下降時，臨界值也會跟著下降，使得回音仍然會被判定為語音訊框。

4.2.2 WFD VAD

以語音樣本的過零量判定結果如圖 16：

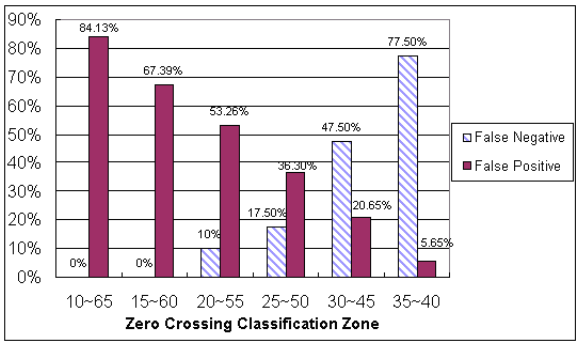

圖 16：以 WFD VAD 演算法分析得到的語音誤判率長條圖

由於無論是正常語音或是回音，都是實際語音，僅有能量的差異，其過零量並沒有顯著的差異，因此顯然無法有效從過零量分析訊框**是否為回音。**

4.2.3 MET VAD 演算法

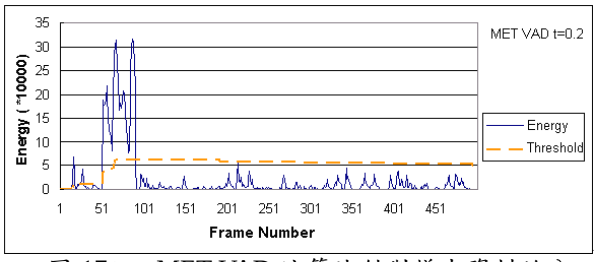

圖 17：以 MET VAD 演算法針對樣本資料的分析結果(t=0.2)

表 3：MET VAD 演算法的誤判率

參數 t	t=0.1	t=0.2	t=0.3
False Positive(%)	7.83	1.74	1.3
False Negative(%)	0	0	5

實驗結果可以看出 MET VAD 演算法只要一開始有語音訊號出現時，即可紀錄此語音訊號峰值作為臨界值的設定依據，隨後即作為辨認之依據。在會談過程中，正常語音與回音的能量變化幅度通常不大，因此所紀錄下的語音峰值用以辨識回音之效果令人滿意。

4.3 實驗二：網路會談實測

本實驗將 MET VAD 分別用於無回音消除機制及有回音消除機制的 VoIP 會談中，測試其針對回音消除的的效果，目的分別為評估 VAD 演算法之 False Positive 以及 False Negative 誤判率。

4.3.1 Flase Positive 誤判率測試

圖 18 為一段包含回音的聲音波形(由 Echo Generator 端麥克風收錄後，未經任何處理)，回音的音量振幅約為正常聲音的 10%。

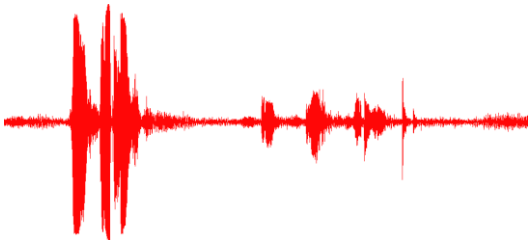

圖 18：含有回音的一段聲音波形

此段聲音波形經過 LED VAD 演算法過濾之後如圖 19：

圖 19：含有回音的聲音波形經過 LED VAD 過濾結果

同樣的聲音經過 MET VAD 後結果如圖 20：

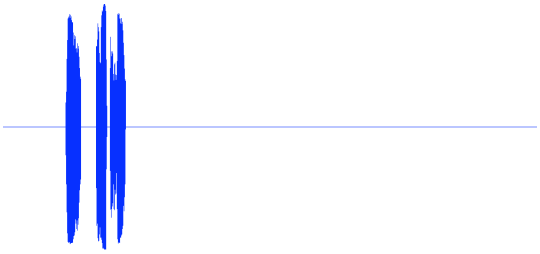

圖 20：含有回音的聲音波形經過 MET VAD 過濾結果

此段聲音在 MET VAD 過濾後，幾乎能夠消除所有非語音訊框，至少 85%以上的回音訊框能成功被消除，僅有與正常語音夾雜的回音無法消除，因此能有效提昇通話品質，使得 MOS 達到 3 以上。除了回音以外，MET VAD 也能夠有效的將音量過低的靜音訊框消除，節省傳輸頻寬同時降低背景雜訊對通話造成的干擾。

4.3.2 Flase Negative 誤判率測試

圖 21 為一段不包含回音的正常語音聲音波形：

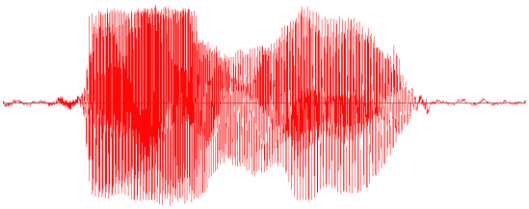

圖 21：不包含回音的語音聲音波形

將此段聲音波形分別輸入 LED VAD 過濾後，其結果輸出波形如圖 22：

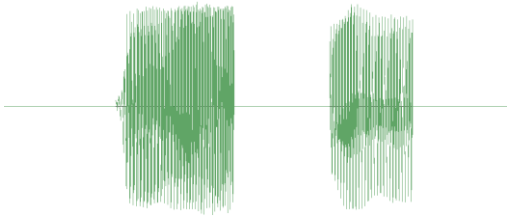

圖 22：不含回音的波形經過 LED VAD 過濾結果

同樣將此段正常語音輸入 MET VAD 過濾，其輸出結果如圖 23：

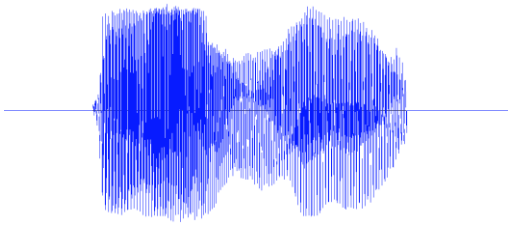

圖 23：不含回音的波形經過 MET VAD 過濾結果

MET VAD 會預先紀錄使用者的說話音量，因此並不容易將語音訊框誤刪，在正常溝通狀況下，誤刪率約在 15%以下，並不會對溝通造成太嚴重的影響。

4.3.3 Skype 通話測試

Skype 發生回音的機率不高，約有 10%的機率無法完全消除回音。目前市面上電腦音效卡均有內建回音消除 DSP (筆記型電腦的麥克風與喇叭位置可預知，更容易預估回音在 acoustic path 所耗時間)，因此回音情況相當罕見。但是一旦回音發生，則所有參與會談使用者的通話品質都會受到影響。

表 4：實驗二結果

演算法	LED VAD	WFD VAD	MET VAD	Skype
% of False Negative	>85%	>90%	<15%	<10%
% of False Positive	>40%	>30%	<15%	<5%

4.4 實驗三：Proximity Problem 的回音消除測試

本實驗針對 Proximity Problem 這種狀況下所產生的回音與雜音，測試 MET VAD 的過濾能力。本實驗採用與 Skype 與自製的 VoIP 程式搭配 MET VAD 分別進行測試。測試時以一對一會談，將兩部電腦距離拉近至 50 公分以內。

4.4.1 測試結果- Skype

使用 Skype 通話時，當兩部電腦距離靠近，互相收到對方喇叭發出的聲音時，當使用者對著麥克風說話後 3 秒之內，即產生 Proximity Problem，造成持續不斷的雜音與回音，其聲音波形如圖 24：

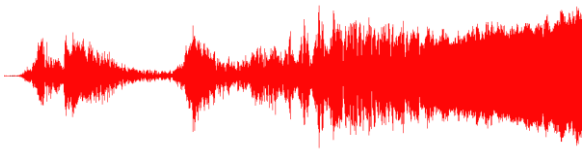

圖 24：Skype 通話時，Proximity Problem 所造成的回音波形

由此證明，Skype 之回音消除機制雖能夠克服大部分回音造成的干擾，但卻無法消除 Proximity Problem 產生的回音。

4.4.2 測試結果- MET VAD

相對的，若使用 MET VAD 過濾，則能夠有效抑制 Proximity Problem 所產生的回音，即使將電腦間距離靠近，也不至於發生回音，此時聲音波形如圖 25：

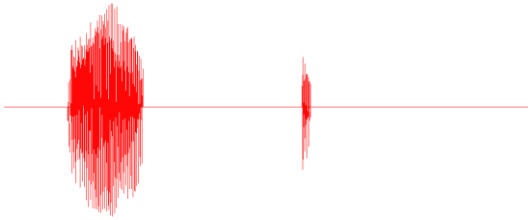

圖 25：MET VAD 有效抑制 Proximity Problem 的結果

五、結論與未來研究方向

本研究分析許多造成現有回音消除機制失效的原因，並指出其發生的狀況與時機，證明在大型語音會談中要徹底消除回音極為困難。因此本研究提出一種以最大語音能量紀錄為基礎的語音動態偵測(VAD)演算法，根據回音的特性，使用最少的運算成本，將回音減至最低。本研究提出的 MET VAD 演算法，比傳統的 VAD 演算法，更能有效的降低回音誤判率，在實驗中，能夠將誤判率降低至 15%以下，有效改善通話品質。若 MET VAD 與現有之回音消除方法配合，將能夠達到互補的作用，且針對 Skype 等軟體所無法克服之 Proximity Problem 所產生的回音，能夠藉由 MET VAD 有效抑制。

除了回音以外，MET VAD 同時能夠有效發揮靜音抑制（Slience Suppression）的效果，阻擋語音會談中不含語音內容的封包。根據實驗，MET VAD 能有效阻擋 85%以上不含語音的靜音封包，且將語音誤刪的機率低於 15%，

有效降低網路頻寬耗用而不影響通話品質。

大型語音會談的另一個影響品質的因素是環境噪音，本研究未來將以環境噪音的消除為主要研究目標。

六、參考文獻

[1] G. S. Fang, "Voice Channel Echo Cancellation", IEEE Communications Magazine, Vol. 21, Issue 9, Dec. 1983, pp.11-14.

[2] Perry P. He, Roman A. Dyba, and Lucio F.C. Pessoa, "Network Echo Cancellers: Requirements, Applications and Solutions", AnalogZONE, 2004.

[3] Brant M. Helf, "Far end echo cancellation method and apparatus", U.S. Patent 4,995,030, Feb. 19, 1991.

[4] M. Hiraguchi, "Full duplex modem having two echo cancellers for a near end echo and a far end echo", U.S. Patent 4,935,919, 19 Jun. 19, 1990.

[5] Yao-Nan Lien, Li-Cheng Chi and Yuh-Sheng Shaw, "A Walkie-Talkie-Like Emergency Communication System for Catastrophic Natural Disasters", Proc. of 10th International Symposium on Pervasive Systems, Algorithms and Networks (ISPAN09), Dec. 14-16, 2009.

[6] B. S. Nollet, and D. L. Jones, "Nonlinear Echo Cancellation For Hands-Free Speakerphones", Proc. of NSIP'97, Michigan USA, Sep. 1997.

[7] G. Periakarruppan, and H. A. Abdul-Rashid, "Packet based echo cancellation for VoIP networks", Computers and Electrical Engineering, Vol. 33, No. 2, 2007, pp. 139-148.

[8] R. V. Prasad, A. Sangwan, H. S. Jamadagni, and M. C. Chiranth, "Comparison of voice activity detection algorithms for voip", Proc. of IEEE Symposium on Computer and Communications, July 2002, pp. 530-535.

[9] R. V. Prasad, R. Muralishhankar, S. Vijay, H. N. Shankar, P. Pawelczak, and I. Miemegeers, "Voice activity detection for VoIP-an information theoretic approach", Proc. of IEEE Global Telecommunications Conference, 2006, pp. 1-6.

[10] P. Renevey, and A. Drygajlo, "Entropy based voice activity detection in very noisy conditions", Proc. of European Conference on Speech Communication and Technology (ISCA EUROSPEECH '01), Sep. 2001, pp. 1887-1890.

[11] Echo suppressor, http://en.wikipedia.org/wiki/Echo_suppressor, Retrieved at November 11, 2009.