

# 一種用於網路電話之遺失封包補償方法

## A Packet Loss Concealment Method for Voice over IP

古鴻炎                      陳佳新  
Hung-Yan Gu              Zia-Sin Chen

國立台灣科技大學資工系  
National Taiwan University of Science and Technology, Taipei, Taiwan  
e-mail: guhy@mail.ntust.edu.tw      http://www.csie.ntust.edu.tw/

### 摘要

本文提出一種用於網路電話的遺失封包補償之方法，稱為 TPPWI 法，它首先使用我們研究的基週偵測方法，來偵測遺失封包的前一個與後一個封包內的語音波形的週期長度，再將前一封包與後一封包各分類為有聲或無聲，依照有聲、無聲之四種組合再分成四種補償作法。其中前、後封包均為有聲的情況，我們提出了基週波形的相位同步方法，需佈放之基週個數與長度的決定方法，執行該二方法後，再執行特別的基週波形內差之作法，如此可得到高品質的重建語音。對於前、後封包均為無聲的情況，本文也將前人的作法加以改進。初步的聽測評估顯示，TPPWI 法可以得到比前人的數種方法更好的語音品質。

關鍵詞：基週偵測，基週波形內差，相位同步，語音品質

### 1. 前言

隨著網路技術的日益蓬勃發展，硬體設備升級、線路頻寬加大，使得透過國際網路來即時傳輸語音封包變得可能，也就是所謂的 VoIP (Voice over IP)。然而，封包交換網路(packet-switching network)本身並不是非絕對可靠的，語音封包在傳輸的過程中，可能被遺失(network loss)，或可能被延遲送達，如果因此而超過了播放的時間點，接收端可能將之丟棄，形成 late loss，這樣的封包遺失(packet loss)對於 VoIP 所要求的對話即時性和語音品質來說，是一個很大的問題。VoIP 接收端的程式，它的工作就是將收到的語音封包解碼為可供播放、聆聽的語音波形，因此當語音封包遺失時，若未作補償，就會解碼輸出一段靜音而形成空隙，使得語音波形變得不連續，因而產生讓聽者可察覺的雜訊噪音。

封包遺失的補償方法(Packet Loss Concealment, 簡稱 PLC)可以分作兩大類[1, 2]，分別是從傳送端(transmitter)與接收端(receiver)的角度進行考量。傳送端補償方法的目標，在於如何降低封包傳送的失敗率與改善其分佈；而接收端補償方法，則致力於提升重建語音的品質。本文所研究的，是一種在接收端進行補償的

方法。

在接收端作補償的方法可細分為：插入式(insertion)、內差式(interpolation)與再生式(regeneration)[1, 2]。插入式方法只會參考過去單邊方向(one-sided)的波形資料，以外差(extrapolation)處理方式來重建語音；內差式方法則會同時考慮遺失封包前後雙邊方向(two-sided)的波形資料，以內差處理方式來重建語音；而再生式方法則是從語音編、解碼器著手，重新合成出遺失封包的語音波形。一般而言，內差式補償方法的運算量較大，但是效果也比較好。考慮到今日個人電腦的速度，時脈(clock rate)都在 1 GHz 以上，要執行內差式的補償方法，運算能力是綽綽有餘，因此我們研究了一種內差式的補償方法，希望讓重建後的語音波形，聆聽時盡可能查覺不出有發生過封包遺失，並且希望在封包遺失率很高時（例如頻寬受限之撥接網路），可以得到大幅度的語音品質改進。

接收端內差式補償方法的原理是，參考遺失發生位置前後封包的資料，來產生兼顧前後連續性的重建語音，雖然步驟比較複雜一些，但是相對的效果也會有顯著的提升。相關的方法包括：**(a)相似波形取代法** [3]，利用樣式比對(pattern matching)的技巧，從過去播放過的波形資料中，尋找可以用來取代遺失封包所造成之空隙的合適波形，由於尋找的不一定都是未遺失的封包，也有可能是重建出來的語音波形，因此會有誤差延續(error propagation)的問題；**(b)基週波形複製法** [4]，此方法仰賴的是有效而可靠的基週偵測演算法，以便重複複製基週波形，直到填滿遺失空隙為止，G.711 [5]內建的封包遺失補償方法便屬於這種；**(c)時域波形伸縮法** [1]，藉由將過去波形伸展拉長，使之能夠涵蓋封包遺失所造成的空隙，進而達到重建語音的效果，WSOLA (Waveform Similarity Overlap Add)波形相似性疊加法便屬於這種作法 [6]，後來經過 Stenger 等人的修改，提出稱為 Modified WSOLA 的作法 [7]，使之能夠作為 VoIP 的接收端補償方法。

接收端補償方法對於遺失封包所重建出的語音信號，其品質好壞的評估可以從如下三項目來檢查，即重建的信號片段與其前後非遺失的信號之間，是否具有**(a)振幅連續性(amplitude continuity)**、**(b)頻率連續性(frequency continuity)**、及邊界上的**(c)相位連續性(phase continuity)**。振幅連續與不連續的對比，可比較圖 1 與圖 2，圖 2 中間重建出的波形，其振幅顯著地比兩邊波形的低；頻率連續與不連續的對比，可比較圖 1 與圖 3；相位連續與不連續的對比，可比較圖 1 與圖 4，圖 4 中間重建出的波形，在右邊界與下一封包交銜接處，出現了相位之不連續。熟知的語音波形重建方法之中，大多數只解決了振幅連續性與封包邊界波形連接的問題，對於頻率與相位連續性的問題考慮不多，或者成效不理想。

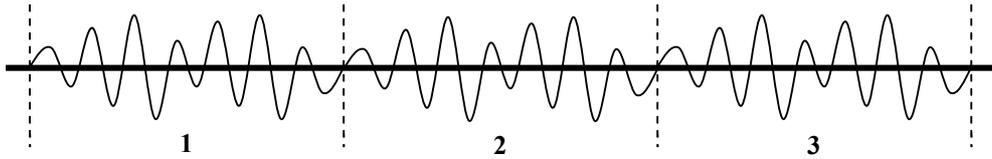


圖 1. 未發生封包遺失的三個連續的語音封包

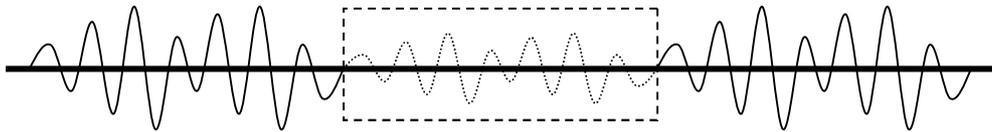


圖 2. 振幅不連續的例子

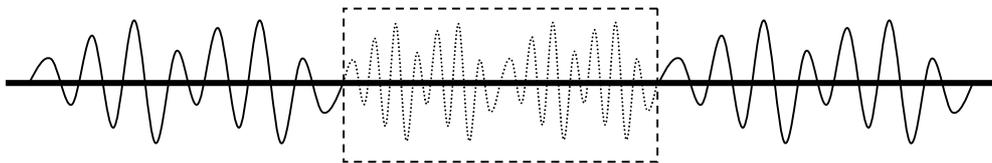


圖 3. 頻率不連續的例子

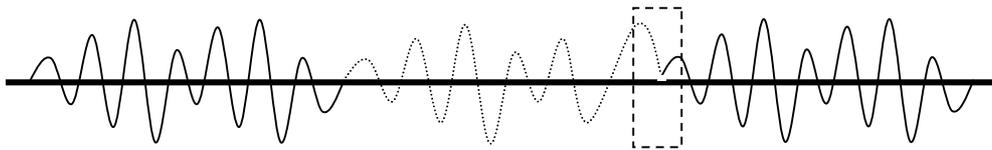


圖 4. 相位不連續的例子

## 2. 補償方法之架構

我們的接收端補償方法之架構如圖 5 所示，先分別從遺失封包之前的那個封包(簡稱“前一封包”)與遺失封包之後的那個封包(簡稱“後一封包”)中，去偵測基週長度、及判斷是否為有聲(voiced)或無聲(unvoiced)的語音，PP (pitch of previous packet)表示前一封包語音的基週長度，PN (pitch of next packet)表示後一封包語音的基週長度；接著，由於前一封包與後一封包的分析為有聲與無聲的結果共有四種組合，所以分別採取不同的波形重建之處理方式，即 BV (both voiced), PV (previous voiced), NV (next voiced), BU (both unvoiced)等處理方式。

雖然如圖 5 所示的架構，並非本文首先提出，而是參考 Liao 等人的雙邊基週波形複製法(Double Sided Pitch Waveform Replication，簡稱 DSPWR)裡的架構 [8]，但是，這樣的架構很明顯地是必需採取的，如此才可能解決頻率與相位之連續性問題。

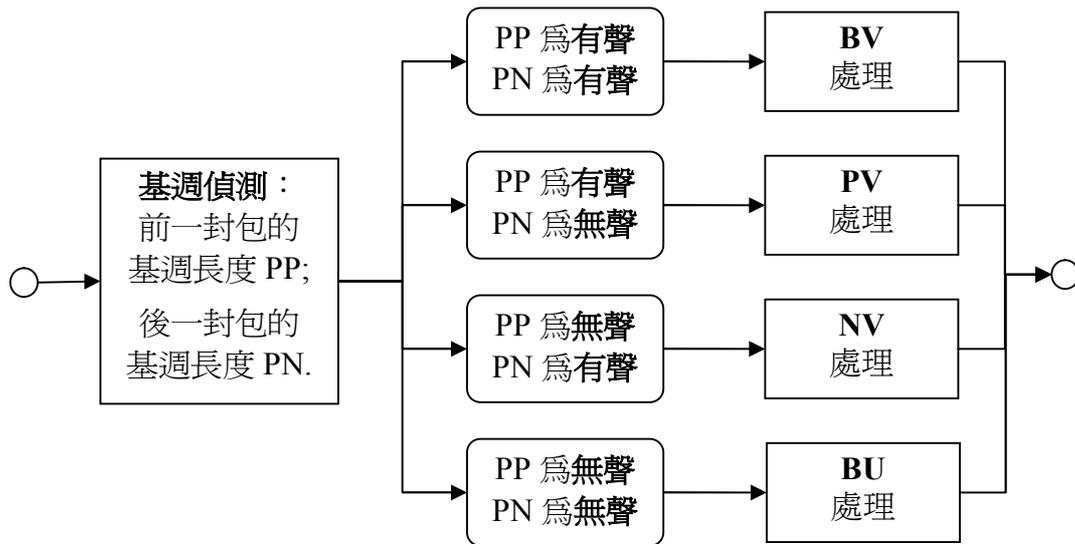


圖 5 補償方法的架構

我們的接收端補償方法稱為時間比例式基週波形內差法(Time Proportion Based Pitch Waveform Interpolation, 簡稱 TPPWI), 雖然它和 DSPWR 法具有相同的架構, 但是細部的處理方法上, 則是南轅北轍且效能也有顯著差異, 如圖 5 裡的”基週偵測”方塊、及右邊最重要的”BV 處理”方塊, 在這兩方塊裡, 我們的方法完全不同於 DSPWR 法裡的, 且效能也比 DSPWR 法的好很多; 在圖 5 裡的”BU 處理”方塊, 我們則是改進 DSPWR 裡的作法, 以得到更好的語音品質; 至於”PV 處理”和”NV 處理”方塊, 我們覺得是較少被執行到而較不重要的方塊, 所以就直接採用 DSPWR 法裡的作法。

接著, 先回顧 DSPWR 法的細部作法, 以說明其缺點所在, 而我們的 TPPWI 法的細部作法, 則在第三、四、五節裡說明。DSPWR 所使用的基週偵測演算法 [9], 是以偵測信號波形的峰點(peak)為基礎, 以峰點間的距離作為基週長度, 此種作法的一個明顯缺點是, 週期性及週期長度的偵測, 準確性不夠高, 這會使得重建出的語音的品質降低不少。此外, DSPWR 法在前後封包均為有聲的情況下, 雖然提出一種解決相位不連續的作法, 但是該作法有一個嚴重的缺點, 就是未考慮音調(pitch)韻律的連續性, 對於聲調語言(如漢語各方言)來說, 音調的表現是重要的。

DSPWR 法在”BV 處理”方塊(PP 與 PN 都是有聲時)的作法如下, 原理是利用相位比對(phase matching)的技巧[10], 來決定 PP 週期個數與 PN 週期個數之線性組合, 以便讓基週波形複製填入到遺失封包所殘留的空隙時, 能夠讓空隙內的重建波形與後一封包之間的波形相位差最小, 藉以舒緩相位不連續的問題, 該程序稱為以基週進行相位比對 (Phase Matching using Pitch, 簡稱 PMP), 最後再使用

線性振幅調整的技巧，來處理振幅連續性的問題。

DSPWR 法在”BU 處理”方塊(PP 與 PN 都是無聲時)的作法，改進了只以前一封包重複的方法[1]，改變成用前一封包的後半段波形與後一封包的前半段波形，來複製填滿遺失封包所產生的空隙，以避免當前一封包裡面包含有聲轉無聲或無聲轉有聲的波形轉變(transition)時，會產生讓人察覺得出來的噪音。

DSPWR 法在”PV 處理”方塊(PP 為有聲而 PN 為無聲時)的作法，原理是以 PP 基週波形來複製填滿遺失封包所產生的空隙，並且對重建波形施行線性振幅調整；由於 PN 被偵測為無週期性，表示後一封包內的波形是隨機跳動之信號，故不存在相位連續性的問題。類似的情況，在”NV 處理”方塊裡，則以 PN 基週波形來複製填滿遺失封包所產生的空隙，並且對重建波形施行線性振幅調整。

### 3. 基週偵測方法及其實驗

#### 3.1 基週偵測

在網路電話的應用裡，許多標準的語音編碼方法都是設定取樣率為 8,000Hz，並且封包的長度最常被設定使用的是 20ms [11]，因此本文針對 20ms (160 個樣本點)的情況研究基週長度的偵測問題，所使用之偵測方法係基於波形相似性的量測，從同一個封包的語音樣本中，選取兩段相鄰且等長的訊號  $s(n)$  與  $s(n + \tau)$  來計算，計算的方式如公式(1)所示 [12]，計算後會得到一正規化自相關 (Normalized Auto-Correlation，簡稱 NAC) 函數，其函數值會介於 1 到 -1 之間。

$$C_{NAC}(\tau) = \frac{\sum_{n=0}^{L-1} s(n)s(n+\tau)}{\sqrt{\sum_{n=0}^{L-1} s(n)^2} \sqrt{\sum_{n=0}^{L-1} s(n+\tau)^2}}, \quad \tau = \tau_{\min}, \dots, \tau_{\max} \quad (1)$$

$$L = \begin{cases} \tau, & \tau_{\min} \leq \tau \leq \frac{N}{2} \\ N - \tau, & \frac{N}{2} < \tau \leq \tau_{\max} \end{cases} \quad (2)$$

其中  $s(n)$  表示語音訊號樣本， $\tau$  表示可能的基週長度之樣本點數， $L$  表示 AC 運算的範圍，其數值設定方式如公式(2)所示， $N$  表示一個封包的語音樣本點數，在  $\tau$  小於等於封包長度之半的情況， $L$  設定為與  $\tau$  的數值一樣，而在  $\tau$  大於封包長度之半的情況下， $L$  的設定則變成封包長度減去  $\tau$  值，例如封包長度  $N$  為 160 個樣本點，則當  $\tau$  計數到 120 時， $L$  將會是 40(即  $160 - 120$ )，亦即計算的是  $s(n)$  與  $s(n + \tau)$  兩段波形訊號的前 40 個樣本點之 AC 係數值。

以一段女性聲音的波形為例(如圖 6 所示)，該波形長度為 20 毫秒具有 160

個樣本點，若該波形為”前一封包”，則正規化自相關函數(NAC)之運算將以由右至左的方向來偵測 PP 長度；若該波形為”後一封包”，則會以由左至右的方向來偵測 PN 長度。以偵測 PP 長度為例，其步驟如下：

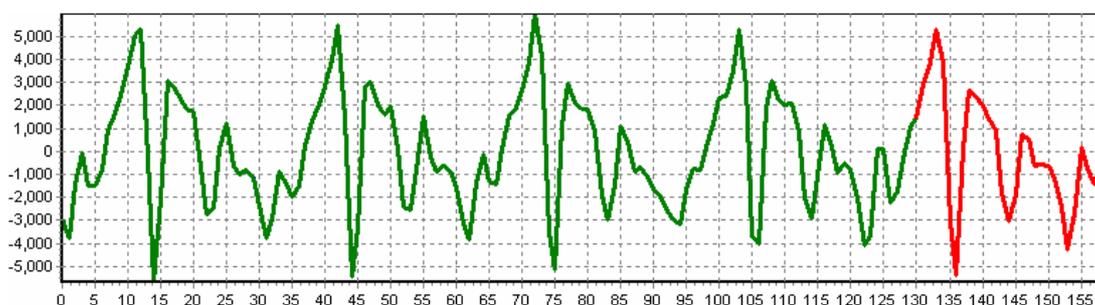


圖 6 女性聲音波形例子

步驟(1.1): 將  $\tau_{\min} - 1$  到  $\tau_{\max} + 1$  範圍內的數值依序代入公式(1)，並將求得之 NAC 係數值連接起來，繪製成如圖 7 所示的函數圖形。

步驟(1.2): 找出  $\tau_{\min}$  到  $\tau_{\max}$  範圍內，NAC 函數的全域最大值之後，將其值乘以 0.8，設定為局部最大值的門檻值 PeakTH，如圖 7 裡所示的水平線。

步驟(1.3): 由左至右對 NAC 函數大於 PeakTH 的數值作檢驗，檢驗其是否為局部最大值(反曲點)，若是則記錄其 X 軸座標(即  $\tau$  值)。以圖 7 為例，局部最大值出現在 30、61 和 91 三個座標點。

步驟(1.4): 令最小  $\tau$  值者為基週長度，驗證其餘的  $\tau$  值是否為其倍數。以圖 7 為例， $\tau$  值最小者為 30，第 2 小者為 61，61 位在可允許的兩倍週期範圍內，亦即介於  $(30 - 5) \times 2 = 50$  到  $(30 + 5) \times 2 = 70$  之間，而且第 3 個  $\tau$  值 91，也介於  $(30 - 5) \times 3 = 75$  到  $(30 + 5) \times 3 = 105$  之間，因此判定 30 為所求的 PP 長度。

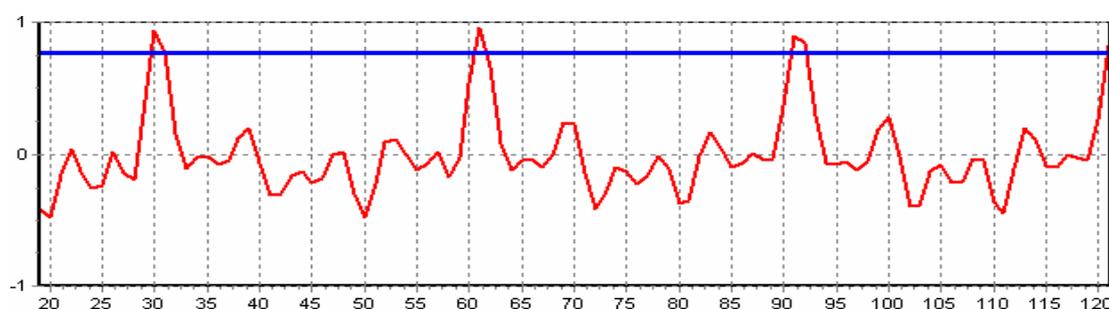


圖 7 NAC 函數圖形

觀察圖 6 的波形，可以確認該封包的 PP 週期長度大致為 30 個樣本點，而整個封包內大約有 5 個這樣的基週波形。

### 3.2 基週確認

在 NAC 函數圖形的 PeakTH 水平線之上，由於女性聲音的基週長度較短，因此女性聲音通常會有多個峰點，而男性聲音的基週長度較長，大多只會有一個峰點。在大於 PeakTH 的峰點只有一個的情況下，為了強化對無聲雜訊與男性有聲聲音的區別能力，本文使用一個波形相似度的判斷：若此峰點的橫座標  $\tau$  值小於等於 50，則此峰點的 NAC 數值必須大於 0.8(波形相似度 80%)，反之，若  $\tau$  值大於 50，因為波形內容的可能變化較多，因此 NAC 係數值只需要大於 0.6(波形相似度 60%)。當 NAC 數值有超過所設定的門檻，才將  $\tau$  值當作基週長度輸出；對於大於 PeakTH 的峰點超過一個的情況下，採用的相似度門檻值則為 0.8。

為了增進基週偵測演算法的正確率，將原先只是偵測前一封包右側與後一封包左側之 PP 與 PN 長度，修改成如下的處理流程：從”前一封包”的右側與左側，分別去偵測左側的基週(PP Left, PPL)與右側的基週(PP Right, PPR)，再從”後一封包”的右側與左側，分別去偵測左側的基週(PN Left, PNL)與右側的基週(PN Right, PNR)；然後，進行基週再確認，即 PPR 與 PPL 比較，PNR 與 PNL 比較，以得到最有可能的 PP 與 PN。

基週再確認的方法，以”前一封包”為例，若 PPR 為 0、而 PPL 為 30，則考慮 PPR 在偵測過程中所繪製的 NAC 函數圖形，檢查  $\tau = 30$  的位置附近是否存在 NAC 係數大於 0.6(波形相似度 60%)的局部最大值；若 PPR 不為 0、而 PPL 為 0 時，也進行類似的處理步驟，藉以更正有聲誤判為無聲的情形。若 PPL 與 PPR 皆不為零，但數值較大者除以數值較小者的比值大於 1.4，則表示兩者數值差異過大，在同一個封包中基週長度不應會發生如此劇烈的變化(例如：PPL 為 45、而 PPR 為 30)，此時就尋找 PPR 的 NAC 函數圖形中  $\tau$  為 45 的附近，是否存在 NAC 係數值大於 0.6 的局部最大值 PPR'，若未找到則令 PPR' 為零；同樣地，在 PPL 的 NAC 函數圖形中  $\tau$  為 30 的附近，找尋 NAC 係數值大於 0.6 的局部最大值 PPL'，若未找到則令 PPL' 為零；接著，計算  $\tau$  值皆約為 30 的 PPR 與 PPL' 的幾何平均數 GM1，以及  $\tau$  值皆約為 45 的 PPL 與 PPR' 的幾何平均數 GM2；若  $GM1 > GM2$ ，則輸出 PPR 之長度 30，若  $GM1 < GM2$ ，則輸出 PPL 之長度 45，藉此更正”週期長度誤判”的情形。

### 3.3 測試實驗

關於基週偵測的測試實驗，我們使用了一些雙字詞的發音語料，其中女性聲音與男性聲音各有十句、各兩組聲音來源，聲調組合包括：下降(44)、加快(14)、右轉(43)、上昇(41)、取消(31)、前進(24)、停止(23)、啓動(34)、清除(12)、關機

(11)。

實驗的方式是，每個封包都會被當作是”前一封包”與”後一封包”，然後分別由右至左與由左至右去偵測 PP 與 PN 的基週長度，所以基週偵測的次數是封包數目的兩倍。程式偵測出 PP 與 PN 數值後，再以人工方式檢查所得的數值的正確性，結果如表 1 所示。其中”誤判週期”表示基週長度被誤判為實際週期的倍數，”誤判有聲”表示無聲信號被誤判為有聲，”誤判無聲”表示有聲信號被誤判為無聲。

表 1 基週偵測方法之正確率

雙字詞 測試語料	封包 數目	偵測 次數	誤判 週期	誤判 有聲	誤判 無聲	正確率
第 1 位女聲	320	640	2	1	8	98.28%
第 2 位女聲	282	564	0	6	1	98.75%
女聲統計	602	1204	2	7	9	98.52%
第 1 位男聲	265	530	1	7	17	95.28%
第 2 位男聲	249	498	16	5	14	92.97%
男聲統計	514	1028	17	12	31	94.12%
整體統計	1116	2232	19	19	40	96.32%

實驗結果顯示，我們的基週偵測方法對於女性聲音的平均偵測正確率為 98.52%，對於男性聲音的平均偵測正確率為 94.12%，整體的平均偵測正確率是 96.32%。男性聲音的基週偵測正確率較低，是因為男聲基週比較長，在使用公式 (1) 作計算時，計算範圍會受到限制。以第 2 位男性聲音為例，他的聲音比較低沉，基週長度經常超過 100 個樣本點，所以他的基週偵測正確率只有 92.97%。

#### 4. 雙邊有聲 BV 之處理

在前、後封包均為有聲(PP 與 PN 均為有聲)的情況下，補償方法的好壞，對重建語音的品質影響是最大的。本文的 TPPWI 法，其主要處理步驟是，先對左、右封包的基週波形取得相位同步並作相位對齊，再對修補後的剩餘空隙訂定需要重建的週期個數、與決定各週期的長度，然後進行基週波形內差處理以產生出各週期的波形內容。這裡的基週波形內差之觀念、作法，是啟發自我們先前研究國語語音合成時提出的 TIPW 信號波形合成法[13]，此外由於 BV 處理是補償方法之重點，因此我們的補償方法的名稱 TPPWI，就取自基週波形內差的縮寫。

##### 4.1 基週波形之同步

由於基週偵測方法只決定基週的長度，因此需要進一步確定一個基週內的波形內容。習用的方法是，將前一封包內最接近遺失封包的 PP 個樣本點，作為前一封包之基週波形(Previous Pitch Waveform, PPW)，而將後一封包內最接近遺失封包的 PN 個樣本點，作為後一封包的基週波形(Next Pitch Waveform, NPW)。但是，本文考慮到下一步驟裡，要對雙邊封包的基週波形作內差，所以必需先對雙邊基週波形的峰點位置(波峰)取得同步(相位同步)，否則當差異太大時(如波峰對應到波谷)，內差得到的振幅將會趨近於零。對於基週波形取得同步的處理，以“前一封包”為例，我們方法的步驟如下：

步驟(2.1): 從前一封包中，選取最右邊、長度為 PP 個樣本點的波形，作為初始的前一基週波形 PPW'，如圖 8 所示。

步驟(2.2): 找到 PPW' 的波峰，以此點為分界，將波形切成左右兩段，再將前半段波形複製到後半段波形的後面，成為前一封包的基週波形 PPW，如圖 9 所示，如此調整後的基週波形之起點將由波峰開始。

步驟(2.3): 比對 PPW' 與 PPW 的相位差，然後在遺失封包的開頭，填入所需的樣本點數，以修補成一個同步過的 PPW 基週波形。圖 8 在作相位修補之後，會變成如圖 10 所示的情況。

至於“後一封包”的 NPW 之選取，也可採取相同之方法，以補滿封包邊界的 PPW 與 NPW 之基週波形，如此就可以解決封包邊界波形平順連接的問題。

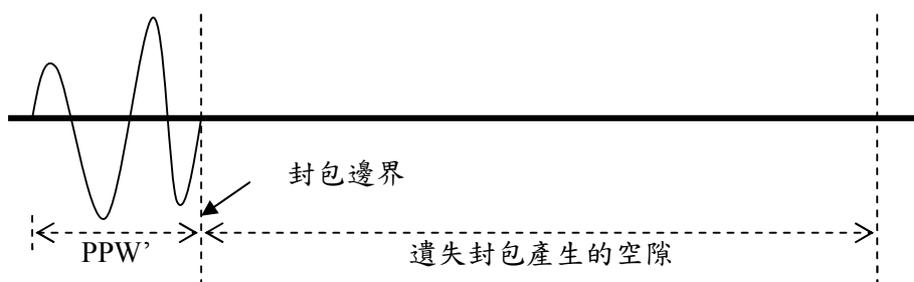


圖 8 前一封包選取出之 PPW' 波形

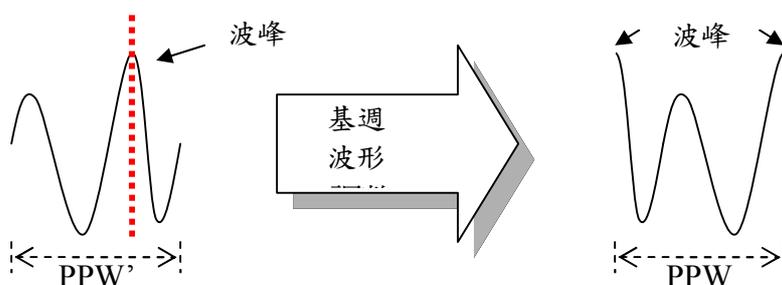


圖 9 從 PPW' 轉換出波峰為起點之 PPW 基週波形

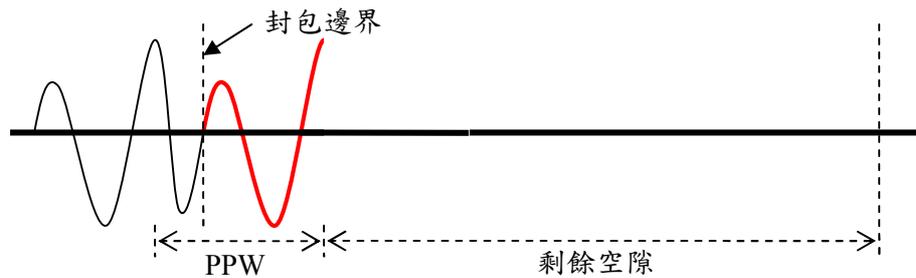


圖 10 補滿封包連接處的基週波形

## 4.2 基週個數與長度訂定

雙邊封包的基週長度關係共有三種情況，亦即  $PP > PN$ ， $PP < PN$ ，或  $PP = PN$ ，因此剩餘空隙內要填入的重建基週波形(Reconstructed Pitch Waveform, RPW)的個數與長度，決定之方法也稍有不同。令剩餘空隙的長度為  $r$ ，首先就依據  $r$  來決定需重建出的基週個數，其步驟如下，將  $r$  除以  $PP$ ，得到  $PPW$  最多可以填入的個數為  $a$ ；將  $r$  除以  $PN$ ，得到  $NPW$  最多可以填入的個數為  $b$ ；取  $a$  與  $b$  的平均再作四捨五入，即為剩餘空隙內要填入的基週個數，令此數值為  $N_p$ 。

接著，以  $PP < PN$  及  $PP=PN$  的情形為例，訂定 RPW 長度的步驟如下：

- 步驟(3.1): 以  $PP$  為基礎，賦予該  $N_p$  個 RPW 的初始長度皆為  $PP$ ，如圖 11 所示。  
將基週長度想像成是”磚塊”數目，如此調整基週長度就相當於調整磚塊的堆積高度。
- 步驟(3.2): 依據斜率，線性地為這  $N_p$  個基週疊上磚塊。以圖 12 為例， $PP$  為 3、 $PN$  為 7、 $N_p$  為 3，故斜率為  $4/3$ ，因此這三個 RPW 的長度就依此斜率來作微調，分別是  $4/3$ 、 $(4/3)*2 = 8/3$  和  $(4/3)*3 = 4$ ，四捨五入後分別為多疊 1 個、3 個與 4 個磚塊。
- 步驟(3.3): 將線性微調後的 RPW 長度相加，得到填入的總長度  $c$ ，以圖 12 為例，此時的總長度為  $4 + 6 + 7 = 17$ 。接著將  $c$  與  $r$  相減，得到差值  $d$ 。若  $d = 0$ ，表示 RPW 的長度在步驟(3.2)的微調處理後，已經是最佳的組合了。
- 步驟(3.4): 若  $d < 0$ ，表示還需要再疊一些磚塊，此時便以由左至右(由低至高)的順序，逐次對一個 RPW 多加 1 個磚塊，直到  $c$  等於  $r$  為止，多加磚塊的順序是第 1 個 RPW、第 2 個 RPW、第 3 個 RPW、回到第 1 個 RPW...
- 步驟(3.5): 若  $d > 0$  的話，則表示需要拿掉一些磚塊，此時便以相反的順序，每次取下 1 個磚塊，直到  $c$  等於  $r$  為止，取下磚塊的順序是第 3 個 RPW、

第 2 個 RPW、第 1 個 RPW、回到第 3 個 RPW...

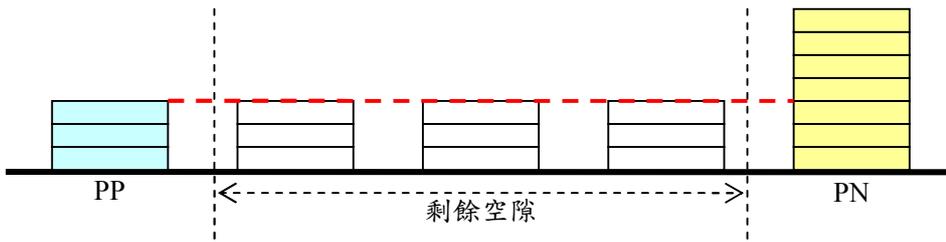


圖 11 填入等長之基週至剩餘空隙

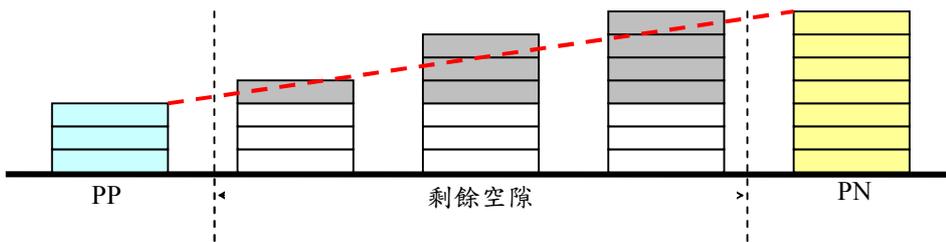


圖 12 線性方式微調後的磚塊數目

至於  $PP > PN$  的情形，亦可採相同之方法，只要將圖 11 的示意圖反轉過來，變成磚塊堆得較高者在左邊即可。

### 4.3 基週波形內差

本文作基週波形內差之方法，步驟如下：

步驟(4.1): 雙邊封包的基週波形 PPW 與 NPW 在 4.1 節便已經決定好了，這裡首先依線性時間比例，來決定它們的內差加權值  $w_1$  與  $w_2$ ，如公式 (3) 所示，

$$w_1 = \frac{r-g}{r}, \quad w_2 = \frac{g}{r} \quad (3)$$

其中  $r$  表示剩餘空隙的長度， $g$  表示要重建的那個 RPW 的中心點位置。例如在圖 13 中，剩餘空隙內只夠填入一個 RPW，因此計算出來的  $w_1 = w_2 = 0.5$ 。另外由公式(3)可知，在剩餘空隙中，離 PPW 較近的 RPW，波形會比較近似於 PPW，相反地，離 NPW 較近的 RPW，其波形會較像 NPW。

步驟(4.2): 分別將 PPW 與 NPW 的樣本點乘以  $w_1$  與  $w_2$ 。例如在圖 13 中，相乘後基週波形的振幅將由 -6000 至 5000 的範圍，降為 -3000 至 2500。

步驟(4.3): PPW、NPW 與 RPW 的長度一般來說都不一樣，此時可使用如公式 (4) 所示的餘弦窗(cos window)，

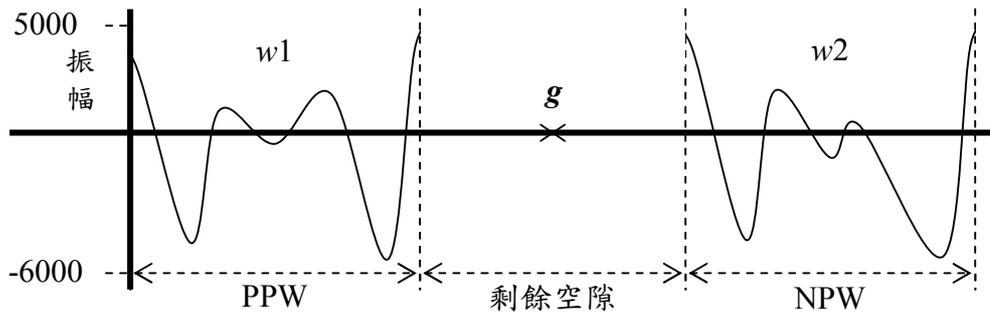
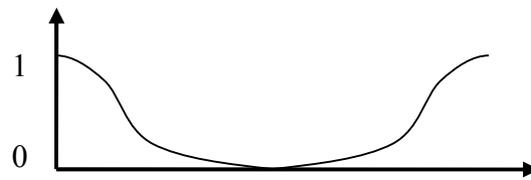


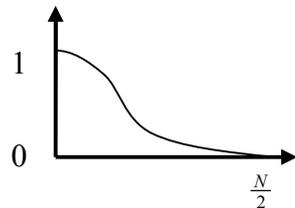
圖 13 加權值  $w_1, w_2$  與重建基週之中心點  $g$

$$w(n) = 0.5 + 0.5 \times \cos\left(\frac{2\pi n}{N}\right), \quad 0 \leq n \leq N-1 \quad (4)$$

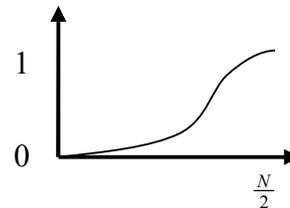
來將 PPW 與 NPW 的波形長度作伸展或壓縮，以便調整成和 RPW 的長度相同。這裡我們不能以 resampling 方式來作，因為單純的 resampling 處理，會破壞原語音的音色[13]。餘弦窗的函數圖形如圖 14 所示，數值介於 0 到 1 之間。



(a) 餘弦窗函數



(b) 左半邊餘弦窗



(c) 右半邊餘弦窗

圖 14 餘弦窗的函數圖形

**伸展處理:** 假設要將長度 30 之 PPW 伸展成長度 35，以配合長度 35 之 RPW，那麼就先設定餘弦窗長度  $N = 30 \times 2 = 60$ ，再將 PPW 的 30 個樣本點乘上左半邊餘弦窗(如圖 14(b)所示)，結果放至 RPW 框的左邊，且與 RPW 框左邊界對齊，如圖 15(a)所示的 PPW-L 波形；接著再將 PPW 乘上右半邊餘弦窗(如圖 14(c)所示)，結果放至 RPW 框的右邊，且與 RPW 框右邊界對齊，如圖 15(b)所示的 PPW-R 波形；然後對重疊的部分作疊加處理。

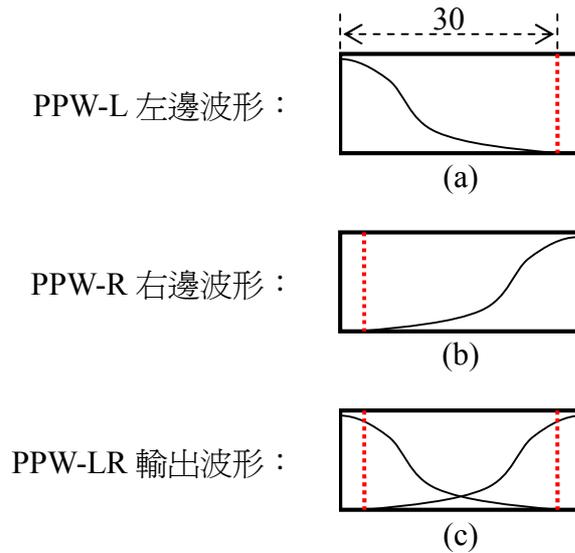


圖 15 以餘弦窗作波形長度伸展

**壓縮處理:** 假設要將長度 30 之 PPW 壓縮成長度 25，以配合長度 25 之 RPW，則先設定餘弦窗長度  $N = 25 \times 2 = 50$ ，再將 PPW 的前面 25 個樣本點乘上左半邊餘弦窗，結果放至 RPW 框的左邊，且與 RPW 框左邊界對齊，如圖 16(a)所示的 PPW-L 波形；接著再將 PPW 的後面 25 個樣本點乘上右半邊餘弦窗，結果放至 RPW 框的右邊，且與 RPW 框右邊界對齊，如圖 16(b)所示的 PPW-R 波形；然後對重疊的部分作疊加處理，而得到 PPW-LR 輸出波形。

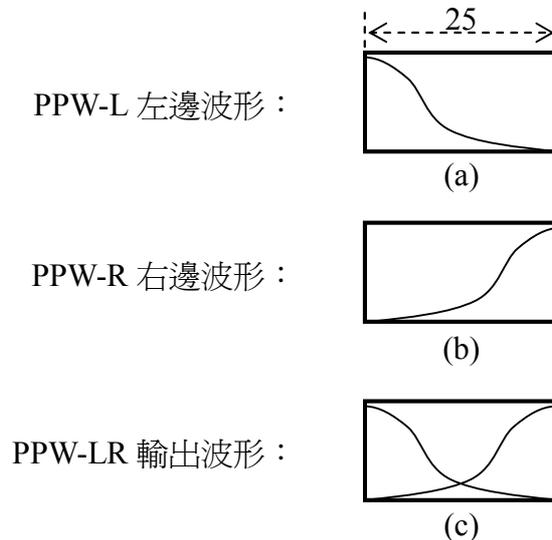


圖 16 以餘弦窗作波形長度壓縮

步驟(4.4): 經過步驟(4.3)的波形伸縮之後，基週波形 PPW、NPW 與 RPW 的長度都變成相同，而且 PPW 與 NPW 的內容在步驟(4.2)已經乘以加權值了，因此只要將它們的每個樣本點相加，即可得到 RPW 波形。圖

17 是將圖 13 的 PPW 與 NPW 基週波形經過振幅加權、長度伸縮與樣本點相加之後，所重建出的 RPW 波形。

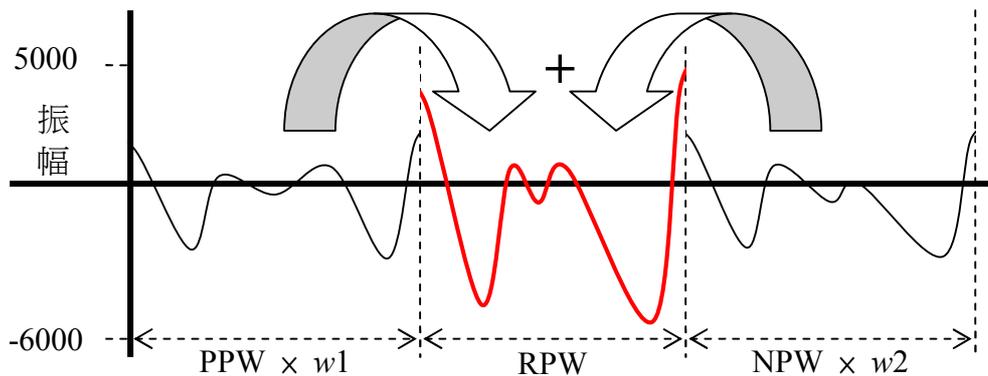


圖 17 重建出的 RPW 波形

## 5. BU、PV 與 NV 之處理

### 5.1 雙邊無聲 BU 之處理

當基週偵測方法在遺失封包雙邊的封包內都偵測不到週期時，表示遺失封包的雙邊封包內的波形是無聲信號，此時就以 BU (Both Unvoiced) 方式處理。以圖 18 為例，若該波形為一個前一封包內的語音波形，隨著時間該波形也由振幅較大且具有週期性的波形，逐漸轉為振幅較小且無週期性的波形，因為在前一封包當中，基週偵測是由右至左進行的，因此會偵測不到語音訊號的週期。為了避免該無聲訊號內可能包含有雜訊的例外狀況，如圖 18 後半段封包的第 2 個四分之一波形所示，需先計算後半段封包的兩個四分之一波形內，各自的峰谷間振幅值；接著，將該二峰谷間振幅值之中數值較大者除以數值較小者，得到一個差異倍數的比值；若該比值大於 1.4，表示可能存在振幅突然變大的雜訊，故只以峰谷間振幅值較小的那段四分之一波形來重複填入兩次到遺失封包內；反之，則表示無聲訊號的變動不大，因此可以拿前一封包的後半段波形來填入遺失封包的空隙內。後一封包亦可以以相同方式來處理。

峰谷間振幅值的差異倍數過大，除了可能存在振幅突起的雜訊之外，基週偵測方法發生誤判也是可能的原因之一，因此透過上述 BU 方式處理，即使基週偵測錯誤，將有聲波形誤判為無聲波形時，仍可減小錯誤波形相位不連續所引發的噪音。

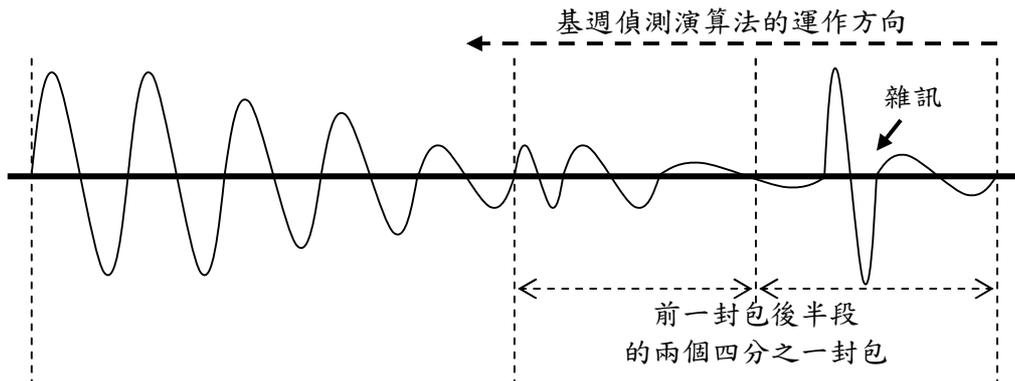


圖 18 週期性逐漸消失的波形例子

## 5.2 單邊有聲 PV 與 NV 之處理

當 PP 為有聲、而 PN 為無聲時，就進行 PV 處理，以 PP 基週波形來複製、填滿遺失封包所產生的空隙，並對重建波形施行前向振幅調整(如圖 19 所示)。

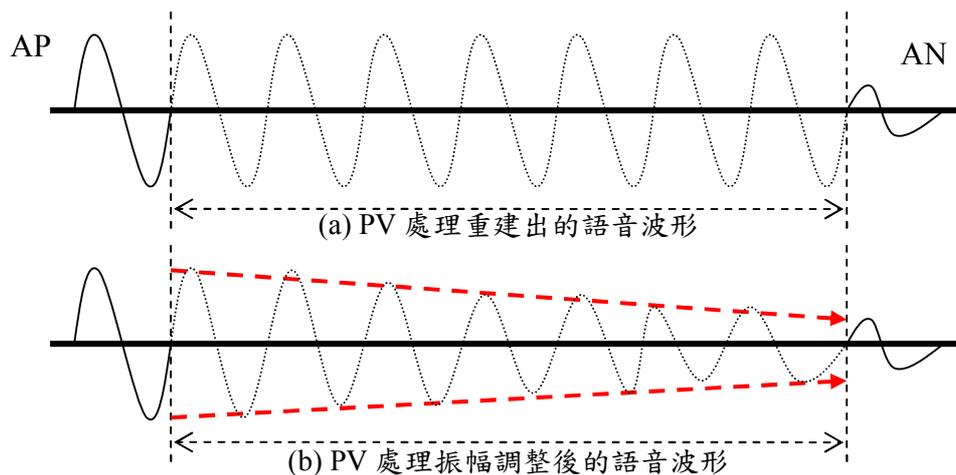


圖 19 前向振幅調整之示意圖

爲了方便說明，我們將前一封包中右邊 PP 個樣本點內的正負峰間振幅值 (peak-to-peak amplitude) 稱爲 AP，而將後一封包左邊 PP 個樣本點內的正負峰間振幅值稱爲 AN。正負峰間振幅值係指一定範圍內，最高波峰與最低波谷之間的差距。重建波形之前向振幅調整，方法如下：

步驟(5.1):  $factor = (AN - AP) / AP / N$ ;

計算振幅需要微調的幅度，其中 AP 與 AN 分別是前一封包與後一封包中，最鄰近重建波形的 PP 個樣本點內，所求得的正負峰間振幅值；N 爲遺失封包空隙的長度，以樣本點計數。如圖 19 裡，後一封包之無聲信號的振幅較小，故 factor 會是負值。

步驟(5.2): for (i = 0; i < N; i++) s[i] \*= 1 + factor \* i;

由左至右逐點調降重建出信號的樣本點振幅值，其中  $s[i]$  表示遺失封包內重建出的信號樣本點序列。圖 19 說明了 PV 處理後，信號振幅逐漸由大變小之情形。

當 PP 為無聲、而 PN 為有聲時，就進行 NV 處理，以 PN 基週波形來複製填滿遺失封包所產生的空隙，其作法相同於 PV 處理的，只是要顛倒左右的方向。

## 6. 補償方法之測試實驗

### 6.1 SNR 訊噪比測試

實驗使用的語音資料，是取自電視連續劇中，兩名演員在餐廳內的對話，取樣率 8000 赫茲、解析度 16 bits、單聲道，時間長度上可以封裝成 20 毫秒為單位的封包共計 2993 個(59.86 秒鐘)。

在模擬封包遺失然後重建其語音波形的實驗中，封包是否遺失，是以隨機亂數搭配白努力(Bernoulli)模型來決定，分成 10%、30%與 50%三種遺失率的情況來模擬。每種遺失率的情況下，我們都作了三次實驗，也就是對同一種封包遺失率模擬產生出三種不同的封包遺失之序列，然後在接收端補償方法處理之後，分成三次計算原始波形與重建波形的 SNR (Signal-to-Noise Ratio)訊噪比，然後取三次計算結果之平均值作為實驗數據。

這裡假設每種接收端補償方法只針對封包連續遺失個數小於等於 3 的情況作處理，理由是：(a)當封包連續遺失個數大於 3 時，重建語音的效果已經明顯不好；(b)同時將多個封包放入 Jitter Buffer 內，可能會影響 Jitter Buffer 的處理效能[11]。因此，除了過去封包重複法之外，其餘補償方法在封包連續遺失個數大於 3 時，便只重複前一次播放過的封包。這裡所作的封包遺失補償之實驗，測試了幾種方法，除了基本的過去封包重複法(Packet Repetition[1]，以 Rep 表示)和本論文提出的 TPPWI 法之外，我們也自行將外差式的相似波形取代法(Waveform Substitution based on Pattern Matching [3]，以 PM 表示)、改良式波形相似性疊加法(Modified Waveform Similarity OverLap Add [7]，以 WSOLA 表示)、與雙邊基週波形複製法(Double Sided Pitch Waveform Replication [8]，以 DSPWR 表示)，寫成可作實驗之程式。

由實驗計算得到的 SNR 值如圖 20 所示，SNR 值的計算公式如下：

$$SNR = 10 \times \log_{10} \frac{\sum_{n=0}^N [x(n)]^2}{\sum_{n=0}^N [x(n) - \tilde{x}(n)]^2} \quad (5)$$

其中  $x(n)$  是原始信號樣本、 $\tilde{x}(n)$  是重建的信號樣本。

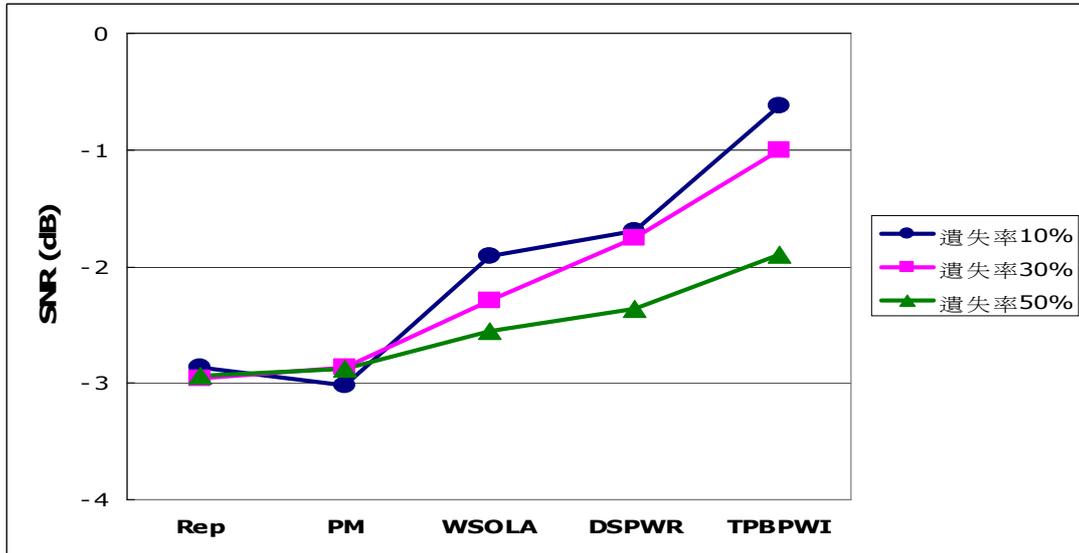


圖 20 重建語音之 SNR

如圖 20 所示的 SNR 的量測結果中，不論在 10%、30%與 50%的封包遺失率之模擬實驗裡，本論文研究之 TPPWI 方法均優於其他四種方法。不過，圖 20 裡重建出的語音波形的 SNR 均為負數，參考 SNR 的計算公式，這表示在  $\log_{10}$  函數內，分母的數值大於分子，使得相除結果為分數，而分母會大於分子，表示此時原始語音的樣本點  $x(n)$  與重建語音的樣本點  $\tilde{x}(n)$  之間的振幅值差異很大，這意謂兩者的波形存在相位差、峰點位置不同等等的情況。

## 6.2 聽測評估測試

針對前述五種接收端補償方法，我們參考 6.1 節的 SNR 測試結果，對它們作排名，第一名到第五名的順序是：TPPWI、DSPWR、WSOLA、Rep 與 PM。接著，我們將這五種方法作配對，第一名與第二名、第一名與第三名、第三名與第四名、第三名與第五名共四種組合，每種組合內包含兩個補償方法處理後的音檔，並且是在模擬封包遺失率為 10%、30%與 50%的情況之下，以相同遺失序列去分別處理得到的，然後讓受測者比較每種組合內的兩個音檔之語音品質差異。評分方式是在受測者不知道組合方式的情況之下，請他們挑選出那一個音檔的語音品質比較好，以及好多少。如果”分不出好壞”的話，就計為 0 分；”好一些”的

話，就計為 1 分；”明顯好”的話，就計為 2 分；”好很多”的話，就計為 3 分。我們總共邀請了 16 位受測者，年齡介於 23 到 25 歲之間，其中 7 位是我們實驗室的成員，另外 9 位是實驗室成員的友人，不具語音專業背景。測試的語料是唐詩楓橋夜泊，先由女生唸一遍、再由男生唸一遍，封裝成 20 毫秒為單位的封包，女性聲音封包數目 535 個、男性聲音封包數目 512 個，總音長 20.94 秒鐘。

聽測後，對 16 位受測者的評分求出平均值，詳細數值如表 2 所示，再依表 2 的數值畫圖，得到圖 21 所示的曲線，從此圖可以看到，在 10% 的封包遺失率之情況下，TPPWI 比第二名的 DSPWR 好了 1.625 分，接近”明顯好”，此時的補償方法效能排序是：WSOLA < PM < Rep < DSPWR < TPPWI；在 30% 的封包遺失率之情況下，TPBPWI 比第二名的 DSPWR 好了 1.375 分，比”好一些”再好一點，此時的補償方法效能排序是：PM < WSOLA < Rep < DSPWR < TPBPWI；在 50% 的封包遺失率之情況下，TPBPWI 比第二名的 DSPWR 好了 0.875 分，接近”好一些”，此時的補償方法效能排序是：WSOLA < PM < Rep < DSPWR < TPPWI。

表 2 聽測實驗之平均分數

封包遺失率	PM	Rep	WSOLA	DSPWR	TPBPWI
10%	0.0625	0.625	0	0.75	2.375
30%	-0.125	0.6875	0	1.0625	2.4375
50%	0.8125	1.3125	0	1.5625	2.4375

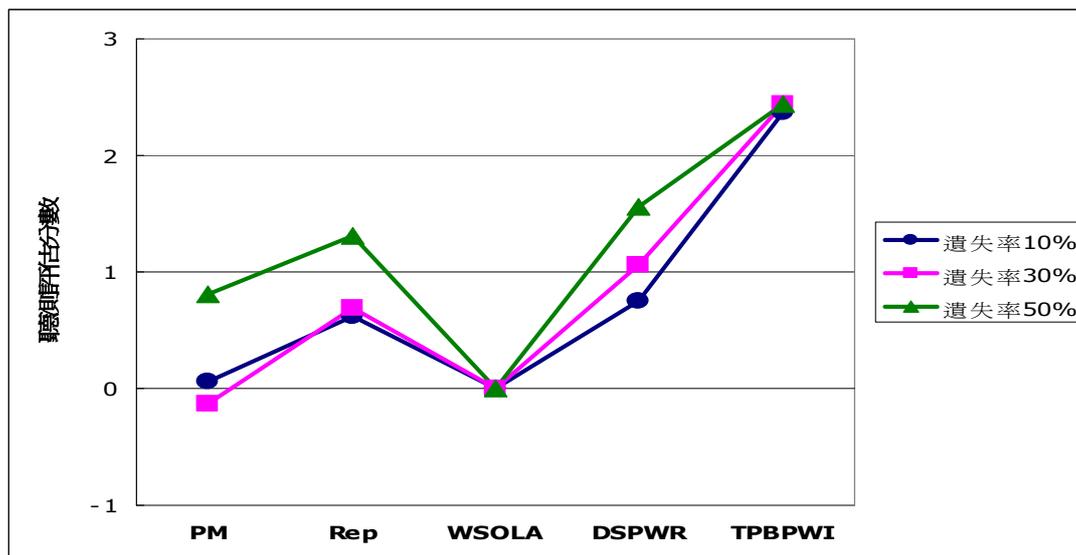


圖 21 聽測實驗之平均評分圖

從這三組數據中我們可以發現，PM 與 WSOLA 的效能比最容易實作的 Rep 略遜一些，原因是它們在搜尋最相似波形的步驟上存在瑕疵，也就是說搜尋範圍

內的波形可能是重建出來的，導致誤差延續，使得它們的補償效果比直接以過去封包重複時還要差。所以 PM 與 WSOLA 兩方法只能夠使用於封包遺失率低於 10% 的情況。另一方面，隨著封包遺失率的升高，DSPWR 與 TPBPWI 之間的效能差距就越近，顯示在越高的封包遺失率之下，接收端補償方法的效果越難發揮，使得各方法彼此之間的補償效能差異變小。

## 7. 結論

網路電話的使用愈來愈普遍，但電腦網路常因繁忙或頻寬不夠而造成封包遺失或延遲，使得通話語音的品質衰退、不穩定，因此研究遺失封包之補償方法，來提升通話語音的品質，是有其重要性的。本文研究了一種接收端之遺失封包補償方法，稱為 TPPWI，它在 SNR 測試和主觀聽測上，都比作比較的 DSPWR、WSOLA、PM、REP 等方法都表現得更好，如在封包遺失率 10% 與 30% 時，聽測評分會比第二名的 DSPWR 分別好 1.625 分與 1.375 分。

此外，本文研究了一種適用網路電話應用的基週偵測方法，量測正規化的自相關函數，週期長度偵測的正確率還不錯，測試實驗顯示可以達到 96.3% 的正確率。在雙邊有聲之 BV 處理，我們提出了不錯的基週波形之同步方法，以及不錯的基週個數與長度的決定方法，再應用先前提出的基週波形內差法，而使得重建出的信號波形具有相位與頻率的連續性，而呈現更高的語音品質。在雙邊無聲之 BU 處理，我們改進了前人的作法，而能夠更有效地避免在無聲信號中，麥克風收錄進來的背景雜訊所引起的噪音。

## 參考文獻

- [1] C. Perkins, O. Hodson and V. Hardman, "A Survey of Packet Loss Recovery Techniques for Streaming Audio", IEEE Network, Vol. 12 (5), pp. 40-48, 1998.
- [2] M. Y. Kim and R. Vafin, "Packet-Loss Recovery Techniques for VoIP", Technical Report, Royal Institute of Technology (KTH), Sweden.
- [3] D. J. Goodman, G. B. Lockhart, O. J. Wasem and W. C. Wong, "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications", IEEE trans. Acoustics, Speech, and Signal Processing, Vol. 34, No. 6, pp. 1440-1448, 1986.
- [4] O. J. Wasem, D. J. Goodman, C. A. Dvorak and H. G. Page, "The Effect of Waveform Substitution on the Quality of PCM Packet Communications", IEEE trans. Acoustics, Speech, and Signal Processing, Vol. 36, No. 3, pp. 342-348,

1988.

- [5] ITU-T, "A High Quality Low-complexity Algorithm for Packet Loss Concealment with G.711", Rec. G. 711 Appendix I, 1999.
- [6] W. Verhelst and M. Roelands, "An Overlap-Add Technique Based on Waveform Similarity (WSOLA) for High Quality Time-Scale Modification of Speech", IEEE ICASSP, Vol. 2, pp. 554-557, 1993.
- [7] A. Stenger, K. B. Younes, R. Reng and B. Girod, "A New Error Concealment Technique for Audio Transmission with Packet Loss", EUSIPCO, 1996.
- [8] W.-T. Liao, J.-C. Chen and M.-S. Chen, "Adaptive Recovery Techniques for Real-Time Audio Streams", IEEE INFOCOM, Vol. 2, pp. 815-823, 2001.
- [9] D. J. Goodman, G. B. Lockhart, O. J. Wasem and W. C. Wong, "Waveform Substitution Techniques for Recovering Missing Speech Segments in Packet Voice Communications", IEEE trans. Acoustics, Speech, and Signal Processing, Vol. 34(6), pp. 1440-1448, 1986.
- [10] R. A. Valenzuela and C. N. Animalu, "A New Voice-Packet Reconstruction Technique", IEEE ICASSP, Vol. 2, pp. 1334-1336, 1989.
- [11] D. Collins, Carrier Grade Voice Over IP, McGraw-Hill Companies, 2000.
- [12] Y. Medan, E. Yair and D. Chazan, "Super Resolution Pitch Determination of Speech Signals", IEEE trans. Signal Processing, Vol. 39(1), pp. 40-48, Jan. 1991.
- [13] Hung-Yan Gu and Wen-Lung Shiu, "A Mandarin-syllable Signal Synthesis Method with Increased Flexibility in Duration, Tone and Timbre Control", Proceedings of the National Science Council, Republic of China, Part A: Physical Science and Engineering, Vol. 22, No. 3, pp. 385-395, 1998.