

利用小波轉換在改善 AMR 語音編碼器之 VAD 與音高週期偵測演算法之研究

Improved VAD and Pitch Detection Algorithms of AMR Speech Codec Based on the Wavelet Transform

吳信德
Hsin-Te Wu

陳嘉祥
Chia-Hsiang Chen
樹德科技大學資訊工程所
高雄縣燕巢鄉橫山村橫山路 59 號
Email:pllo0304@mail2000.com.tw

陳璽煌
Shi-Huang Chen

摘要

本文針對適應性多重位元率(Adaptive Multi Rate, 簡稱 AMR)語音編碼器提出一套可同時達成 VAD(Voice Active Detection)與 Pitch Detection 的新型演算法,該演算法先利用 VAD 所求得有聲部分來進一步計算 Pitch 週期,本文提出的 VAD 演算法技術以知覺小波封包轉換(Perceptual Wavelet Packet Transform, 簡稱 PWPT)與適應性加權門檻值(Adaptive Weighted Threshold, 簡稱 AWT)為基礎來辨識語音訊號中無聲部分與有聲部分,由於語音訊號大部分都含有背景雜訊,所以將辨識出的有聲段經由 FNBS(Filter Noise Band Select)將雜訊濾除,最後將有聲部分透過空間域自相關函數與音高週期校準後,便可以找出語音訊號的 Pitch 週期,從實驗結果可得知本文提出 VAD 演算法優於 AMR 編碼器中的 VAD,並且可以濾除有聲段語音訊號雜訊,同時配合 VAD 的 Pitch Detection 演算法其運算時間較原先 AMR 編碼器快,並可降低原先 AMR Pitch Detection 處理的複雜度。

Abstract

This paper proposes improved voice active detection (VAD) and pitch detection algorithms for AMR speech codec. The proposed VAD algorithm uses perceptual wavelet packet transform and adaptive weighted thresholding to detect the voiced and unvoiced frames. Due to speech signals usually contain background noises, the detected voiced

frames will be denoised via filter noise band select method. Then, the pitch information can be derived from the denoised voiced frames by the use of spatial correlation function and pitch correction scheme. It follows from experimental results that the proposed algorithms are superior to the recommended VAD and pitch detection algorithms in AMR standard. In addition, the proposed algorithm can reduce the complexity of VAD as well as pitch detection in AMR speech codec.

關鍵字: AMR、VAD、小波轉換、音高週期偵測。

Keywords: AMR、VAD、Wavelet Transform、Pitch Detection。

1. 前言

AMR 編碼器是由第三代合作計畫(3GPP)組織所提出[1-2],這個組織的目標是推動並訂定下一代無線通訊系統的相關標準與規格,而 AMR 是以『代數碼簿激發線性預測』(Algebraic Code Excited Linear Prediction, ACELP)技術為設計基礎,並且總共有八種編碼位元率,分別為 4.75、5.15、5.90、6.70、7.40、7.95、10.2 以及 12.2 Kbps, AMR 可依頻寬及網路壅塞情況選擇最佳的編碼模式。AMR 為了提高語音壓縮率並同時提升網路的傳輸功率,其編碼器使用 VAD 處理,AMR 中的 VAD 模組是用來辨識語音訊號中的有聲部分與無聲部分,編碼器只需要傳送有聲段部份,因此可提升語

音訊號壓縮率。

常見的 VAD 演算法中通常使用下列的偵測門檻特徵值，例如：Short Time Energy Levels、Zero-Crossing Rates 或 Pitch Period，這些方法都限制使用在時域上，並且一段語音訊號通常是不規則性且包含著瞬間變化訊號所構成的，因此在附加雜訊語音環境中僅使用時域判斷擷取 VAD 特徵的方法是不恰當的，尤其，AMR 中的 VAD 使用預設的門檻值在時域上判斷，雖然易於實現在編碼器上，但是在附加雜訊語音環境下容易造成誤判且效率不佳[4-7]。

ACELP 其架構有一組閉迴路系統用以搜尋最佳適應性碼簿(Adaptive Codebook)以及固定碼簿(Fixed, Algebraic or Innovation Codebook)[3]，適應性碼簿用來代表有聲段信號特徵如音高週期，而固定碼簿則是用來代表無聲段信號特徵如高斯雜訊信號，由於適應性碼簿所代表的音高週期特徵為人類語音信號中最重要基頻成分，所以 AMR 語音編碼器另外對音高週期加了一組開迴路搜尋機制，搭配適應性碼簿閉迴路搜尋用以提升音高週期搜尋的準確性，因此音高週期最佳化的搜尋動作是 AMR 編碼器耗時較多的步驟之一，現今常見的 Pitch Detection 演算法包括自相關函數、平均振幅差函數...等，而 AMR 中 Pitch Detection 使用自相關函數尋找每一個子音框(每 40 樣本為一子音框)的音高週期，最後透過閉迴路校準音高週期，自相關函數音高週期偵測演算法在乾淨語音環境下效果不錯，但在附加雜訊語音環境下容易受到干擾，並且一個子音框可能有一個以上的 Pitch 週期，所以必須仰賴大量校準機制來做修正。

為了克服上述問題，本文提出一種可同時達成 VAD 與 Pitch Detection 的新型演算法來提升 VAD 與 Pitch Detection 在 AMR 中的效能，本文提出的 VAD 演算法使用 PWPT[1]與 TEO (Teager Energy Operator，強化能量運算) [2]來提供系統一個可變的時間頻率對應關係，並且產生一個頻率和時間變化的能量曲線，最後透過 AWT 計算提高 VAD 門檻值精確度，最後將辨識出的有聲部分經由 FBNS 濾除有聲部分的附加雜訊，降低附加雜訊

影響 Pitch Detection 的準確度，再經由本文提出 Pitch Detection 找尋正確的音高週期，同時本文提出 Pitch Detection 演算法可以在附加雜訊語音環境下正確找出音高週期，而本文提出的 VAD 演算法不需事先預設門檻值，而是透過各種不同語音環境計算取得，由實驗結果可得知本文提出 VAD 與 Pitch Detection 確實優於 AMR 預設模組。

2. 改良式 VAD 模組

2.1. 知覺小波封波轉換

本文提出的 PWPT[8-9]可將輸入訊號分解成 17 個子波段訊號，其分解樹狀階層共有五階，使用的語音訊號取樣率設定為 8 kHz，根據奈奎斯(Nyquist)取樣定理所以頻寬設定在 4 kHz，在人耳聽覺範圍 0~16 kHz 可以分成 24 個臨界頻帶，由於語音訊號所能分佈的頻率大約在 0~4 kHz，所以只要根據 0~4 kHz 的 17 個臨界頻帶範圍分解訊號，表 1 為 17 個子波段訊號頻率範圍[10]，而表 1 中 CBW(Corresponding Critical Bandwidth)為中間頻率，其計算公式為公式(1)。

$$CBW(fc) = 25 + 75(1 + 1.4 \times 10^{-6} fc^2)^{0.69} \text{ Hz} \quad (1)$$

PWPT 之設計模式須匹配人耳的聽覺模型，此作法可以改進各種基於小波轉換的語音處理系統，PWPT 分解輸入訊號的方法是使用小波轉換 Daubechies 中的 db5 方法，圖 1(a)為小波轉換分解細胞(Cell)，圖 1(b)為反向小波轉換合成細胞，其中 $c_1(k)$ 為近似(Approximations)係數也就是訊號低頻部份， $d_1(k)$ 為細節(Detail)係數也就是訊號中高频部份，而 h_0 與 h_1 分別為分析低通及高通濾波器， $g_0(n)$ 與 $g_1(n)$ 則分別為合成低通及高通濾波器，同時符號 $\downarrow 2$ 及 $\uparrow 2$ 分別代表兩倍減少取樣(Down-sampling by 2)以及兩倍增加取樣(Up-sampling by 2)，而 $x(n)$ 為原始訊號，PWPT 分解樹是由 16 個基本小波轉換分解細胞所組成，公式(2)為低頻分解公式，公式(3)為高频分解公式，我們透過上述觀念就可以作多層的分解，並且也可以建立出 PWPT 樹狀結構如圖 2 所示。

表 1、知覺小波封包轉換 17 個子波段

No.	$f_c(\text{Hz})$	CBW	No.	$f_c(\text{Hz})$	CBW
1	50	-	10	1170	190
2	150	100	11	1370	210
3	250	100	12	1600	240
4	350	100	13	1850	280
5	450	110	14	2150	320
6	570	120	15	2500	380
7	700	140	16	2900	450
8	840	150	17	3400	550
9	1000	160			

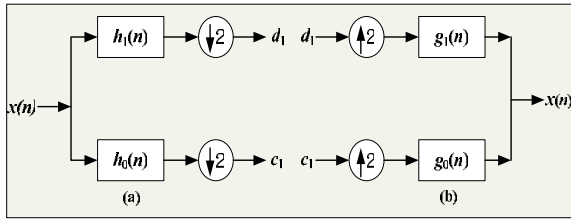


圖 1、小波轉換分解過程與反向小波轉換過程

$$c_1(k) = \sum_n h_0(n-2k)x(n) \quad (2)$$

$$d_1(k) = \sum_n h_1(n-2k)x(n) \quad (3)$$

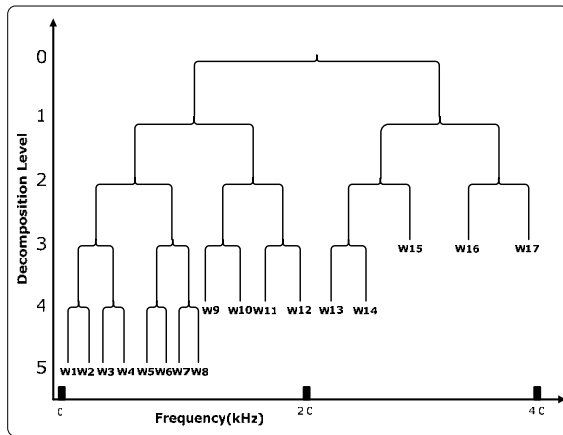


圖 2、知覺小波封包轉換樹狀圖

2.2. 強化能量運算

由於有聲部分是屬於穩定或半穩定訊號，而無聲部分是屬於不穩定訊號，TEO[11-12]的功用是強化穩定或半穩定訊號，並衰減不穩定訊號，TEO 是一個極為有效、非線性計算的演算法，透過機械

和物理考量能提取信號能量，公式(4)為 TEO 計算公式， y 為 PWPT 所建立的 1~17 個子波段訊號， k 為子波段訊號的長度， $T(y[k])$ 是子波段訊號經由 TEO 處理後的結果。

$$T(y[k]) = y[k]^2 - y[k+1]y[k-1] \quad (4)$$

2.3. 波段選擇

由於透過 TEO 計算後，可強化有聲部分並衰減無聲部分，子波段訊號可以透過波段選擇(Band Selection)計算將有可能是有聲部分的波段作保留，確定是無聲部分的子波段訊號可使用波段選擇計算公式(5)[4]將雜訊濾除， N 為每一階層長度， $t_{j,m}(i)$ 為經由 TEO 計算後的子波段訊號， $\sigma_{j,m}$ 之計算公式如 (6) 所示，而 MAD(Menu Absolute Deviation)為平均絕對值誤差， $MAD_{j,m}$ 的作法是將整個子波段取得平均值 Q ，再將平均值 Q 與整個子波段每點相減後取絕對值再將誤差相加就可得到 $MAD_{j,m}$ ，公式(7)是用來判斷子波段是否為無聲部分或有聲部分，其中 $var\{t_{j,m}(k)\}$ 是取得整個子波段的變異數，若 $var\{t_{j,m}(k)\}$ 大於 $\lambda_{j,m}$ 表示為有聲部分則保留，若小於 $\lambda_{j,m}$ 則表示為無聲部分可將整個子波段訊號歸零取代。

$$\lambda_{j,m} = \frac{1}{k} \sum_{i=1}^k \frac{\sigma_{j,m} \sqrt{2 \log(N)}}{t_{j,m}(i)} \quad (5)$$

$$\sigma_{j,m} = \frac{MAD_{j,m}}{0.6745} \quad (6)$$

$$T_{j,m}(k) = \begin{cases} t_{j,m}(k), & \text{if } var\{t_{j,m}(k)\} \geq \lambda_{j,m} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

2.4. 聲音活動狀態曲線

聲音活動狀態曲線(Voice Activity Shape, VAS)計算是為了提供 VAS 給 AWT 計算門檻值使用，VAS 是從 IPWPT 計算後的這些子波段訊號計算取得，VAS 計算公式如(8)所示， $RW_{j,m}(k)$ 為 IPWPT 計算後的子波段訊號，VAS 計算作法是將 IPWPT 計算後的子波段訊號全部相加就可以得到 VAS。

$$V(n) = \sum_{m=1}^{17} RW_{j,m}(k) \quad (8)$$

而由於VAS波形中鋸齒狀波形太多，而這些鋸齒狀波形會影響門檻值的精準度，所以VAS經由滑動平均濾波器(Running Average Filtering)將鋸齒狀波形平滑化後，就可得到 $V'(n)$ ，滑動平均濾波器公式如(9)所示，其中 L 為一參數，此處 $L=32$ 。

$$V'(n) = \frac{1}{L} \sum_{i=0}^{L-1} V(n-i), n=1,2,\dots,N \quad (9)$$

2.5、適應性加權門檻值

當求出VAS時可以透過適應性加權門檻值來得到一個強健的門檻值，圖3為AWT實驗結果，而適應性加權門檻值[13]計算步驟如下：

- (1) 初始化變數令 $k=1$ ，並且 $V^{(1)}(n) = V'(n)$ 。
- (2) 令 $V^{(k+1)}(n)$ 定義公式如下：

$$V^{(k+1)}(n) = \begin{cases} V^{(k)}(n), & \text{if } V^{(k)}(n) < E[V^{(k)}(n)] \\ E[V^{(k)}(n)], & \text{otherwise} \end{cases} \quad (10)$$

(10)

$E[V^{(k)}(n)]$ 為 $V^{(k)}(n)$ 的平均值。

- (3) 重複步驟2，就可從 $E[V^{(2)}(n)]$ 中得到二次遞迴平均值(second derivative round mean, SDRM)如圖3所示。
- (4) 定義 p 為加權值， p 的加權值計算為 $p = Lv / L$ ，其中 Lv 為 $V^{(2)}(n)$ 的能量和，而 L 為輸入訊號的能量和。
- (5) 最後，AWT的結果可從 $AWT = p \times \{E[V^{(2)}(n)] + E[V^{(3)}(n)]\} / 2$ 取得。

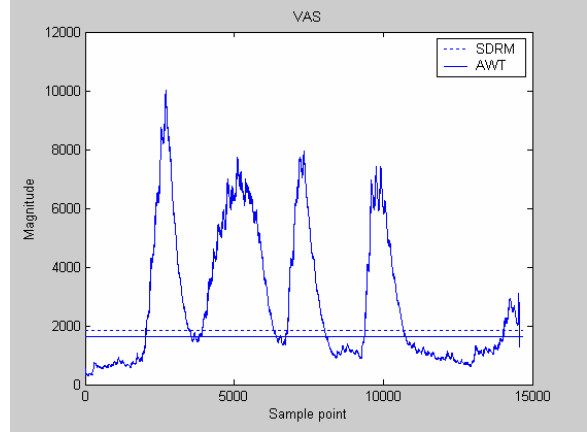


圖3、AWT實驗結果

2.6. 濾除雜訊波段選擇

當PWPT將訊號分解成17個子波段後，將17個子波段經由FNBS計算把每一波段的雜訊濾除，濾除雜訊門檻值計算公式為(11)是用來計算每一個波段每一點門檻值， γ_j 計算公式為(12)是用來計算每一個波段門檻值， $M'_{j,m}(k)$ 計算公式為(13)，其中 $w_{j,m}(k)$ 為PWPT分解的子波段訊號，計算公式(14)為判斷每一個波段每一點是雜訊還是乾淨語音，判斷子波段每一點是否小於 $\gamma_{j,m}(k)$ 如果是就以 $H_{j,m}(k)$ 取代，而 $H_{j,m}(k)$ 計算公式(15)，雖然判斷小於 $\gamma_{j,m}(k)$ 但訊號中還是會有參雜少量雜訊能量所以透過 $H_{j,m}(k)$ 計算將少量雜訊濾除，sgn代表訊號的正負號。

$$\gamma_{j,m}(k) = \gamma_j (1 - M'_{j,m}(k)) \quad (11)$$

$$\gamma_j = \sigma_{j,m} \sqrt{2 \log N} \quad (12)$$

$$M'_{j,m}(k) = \left[\frac{w_{j,m}(k)}{\max(w_{j,m}(k))} \right] \quad (13)$$

$$S'(n) = \begin{cases} H_{j,m}(k), & \text{if } |w_{j,m}(k)| < \gamma_{j,m}(k) \\ \text{sgn}(w_{j,m}(k)) (|w_{j,m}(k)| - \gamma_{j,m}(k)), & \\ \text{otherwise} \end{cases} \quad (14)$$

$$H_{j,m}(k) = \text{sgn} \left(\left(1 + w_{j,m}(k) \right)^{\frac{|w_{j,m}(k)|}{\gamma_{j,m}(k)}} - 1 \right) \quad (15)$$

3. Pitch Detection

3.1. 空間域自相關函數

透過 VAD 辨識的有聲部份以及除雜訊後的語音訊號當作空間域自相關函數的輸入訊號，由於大部分的語音訊號都含有背景雜訊，這些雜訊往往會影響找 Pitch 週期的正確性，以 AMR 預設的 Pitch Detection 採用的是時域自相關函數方法，若要在時域上濾除背景雜訊並找尋正確的 Pitch 週期是不容易的，所以本文提出 Pitch Detection 使用在頻域上，並且使用以小波轉換為基礎的空間域自相關函數來抵抗語音訊號中的附加雜訊。

空間域自相關函數[14-16]作法首先將 VAD 以及除雜訊後的語音訊號透過小波轉換分解二階的樹狀結構，小波轉換是使用 Daubechies 中的 db5 方法，利用 0~2kHz 與 0~1kHz 的低頻波段透過反向小波轉換將兩個波段恢復到原來長度，再將兩個波段相乘就可以得到空間域自相關函數，空間域相關函數計算公式(16)， l 為一參數，此處 $l=2$ 。

$$Corr_1(m, k) = \prod_{j=0}^{l-1} RW_{j, m}(k) \quad (16)$$

3.2. Pitch 週期校準

空間域相關函數原先是用來消除語音訊號雜訊，使用在 Pitch Detection 上可進一步加強其抗雜訊的能力，音高週期校正首先利用空間域相關函數進行正規化後可以得 $\{resCorr_2(m, k)\}$ 其計算公式(17)，其中 $Pa(k)$ 計算公式(18)與 $PCorr(k)$ 計算公式(19)，最後用來求取音高週期的新空間域相關函數 $\{NewCorr_2(m, k)\}$ 由計算公式(20)得到，其中 $w(k)$ 為一個長度為 $[30ms \times Fs]$ 的 FIR 低通濾波器， Fs 為信號取樣頻率。

$$resCorr_2(m, k) = Corr_1(m, k) \sqrt{Pa(k) / PCorr(k)} \quad (17)$$

$$Pa(k) = \sum_k \{RW_{j, m}(k)\}^2 \quad (18)$$

$$PCorr(k) = \sum_k \{Corr_1(m, k)\}^2 \quad (19)$$

$$NewCorr_2(m, k) = resCorr_2(m, k) * w(k) \quad (20)$$

本文利用兩種音高週期搜尋方法來找尋正確的音高週期，第一種方法首先求取音高週期特徵，而求取的準則為當 $|NewCorr_2(n)| \geq |RW_{j, m}(k)|$ 時，位址指標 n 便可視為音高週期特徵發生點，並將其紀錄在向量 $G(n)$ 中，接下來做音高週期間隔作判斷，由於音高週期間隔不會瞬間變化很大，所以首先可對 $G(n)$ 內音高特徵間隔大於平均音高週期 2 倍的元素做移除，而小於平均音高週期 1/2 倍的元素與最鄰近的點比較能量，能量大的保留，能量小的移除，再將 $G(n)$ 內記錄的位址在空間域相關函數上做音高週期位址校正，完成前述校正動作後就可以得到頻譜自相關函數大約分佈的音高週期，為了得到更正確的音高週期再利用第二種方式來做輔助校正。

第二種方式利用事先求取得到的 $Corr_1(m, k)$ 取音框(Frame)，利用計算公式(21)得到音高週期分佈門檻值 $\{PitchTr\}$ ， $Max(n)$ 與 $Ave(n)$ 分別為音框的最大值與平均值，將音框中訊號大於 $PitchTr$ 的訊號保留，小於 $PitchTr$ 的訊號移除就可以得到音框中音高週期分佈區域，再從音高週期分佈區域中搜尋最大值位址紀錄在向量 $G1(n)$ 中，而位址指標 m 便可視為音高週期特徵發生點，再透過 $G1(n)$ 的紀錄位址在空間域相關函數上做音高週期位址校正，此時 $G1(n)$ 的紀錄位址合併 $G(n)$ 中，最後 $G(n)$ 內的元素出現的平均週期便是該語音段精確的音高週期搜尋值。

$$PitchTr = Max(n) - ((Max(n) - Ave(n)) \times 0.6) \quad (21)$$

4. 實驗結果

本篇論文中所使用的硬體設備為 Pentium4 1.6G，程式開發軟體為 Visual C++ 6.0、MATLAB 7.0，測試語音訊號來源為奧羅拉英文標準音檔，圖 5 與 6 所使用的範例是 FAP_8986A 訊號雜訊比為 5dB，圖 4 所示上圖為 VAS 曲線下圖為原始音檔，圖 5 為本文提出 VAD 方法實驗結果，上圖為 VAD 判斷後的結果下圖為原始音檔，由上述得知經由本文提出方式可以得到較為精準的 VAD 門檻值並可將大部分雜訊濾除。

表 2 實驗結果是由本文提出的 VAD 演算法與 AMR 的 VAD 模組實驗後的比較，語音訊號來源是從奧羅拉英文標準音檔中 20dB~0dB 與無雜訊亂數選取共 1200 個音檔來當做實驗範例，由表 2 可以得知本文所提出的方法優於 AMR 的 VAD 模組，而本文提出的方法在任何語音環境都有不錯的效率，表中 Pf 代表為語音訊號有聲段經由 VAD 判斷為無聲段，圖 6 為本文與其他 VAD 演算法在 Pf 效能上的比較， Pd 代表為語音訊號有聲段經由 VAD 判斷為有聲段，圖 7 為本文與其他 VAD 演算法在 Pd 效能上的比較， Nf 代表為語音訊號無聲段經由 VAD 判斷為有聲段，圖 8 為本文與其他 VAD 演算法在 Nf 效能上的比較， Nd 代表為語音訊號無聲段經由 VAD 判斷為無聲段，圖 9 為本文與其他 VAD 演算法在 Nd 效能上的比較。

透過本文提出 VAD 演算法判斷且濾除雜訊後的有聲部分當作 Pitch Detection 輸入訊號，由圖 10、11 所示可以得知透過 VAD 搭配 Pitch Detection 確實可以正確找出語音訊號中的音高週期，圖 10 與圖 11 所使用的範例是 Nine 與 One 音檔訊號雜訊比為 10dB，圖 10 與圖 11 中 (a) 為原始音檔、(b) 為原始音檔中的 1000~1700 點原始訊號、(c) 是經由本文提出的 Pitch Detection 演算法所偵測出 1000~1700 點中的音高週期，表 3、Pitch Detection 實驗結果中的語音訊號來源是從奧羅拉英文標準音檔中 20dB~0dB 與無雜訊男女聲各亂數選取共 1200 個音檔來當做實驗範例，圖 12 為本文與 AMR 中的 Pitch Detection 模組平均校正次數(男聲)比較，圖 13 為本文與 AMR 中的 Pitch Detection 模組平均校正次數(女聲)比較，圖 14 為圖 12 與圖 13 平均誤差(男聲、女聲)比較，由表 3 可以得知本文提出的方式確實優於 AMR 預設的 Pitch Detection 模組，並且可以適用於各種附加雜訊語音環境下得到不錯的結果。

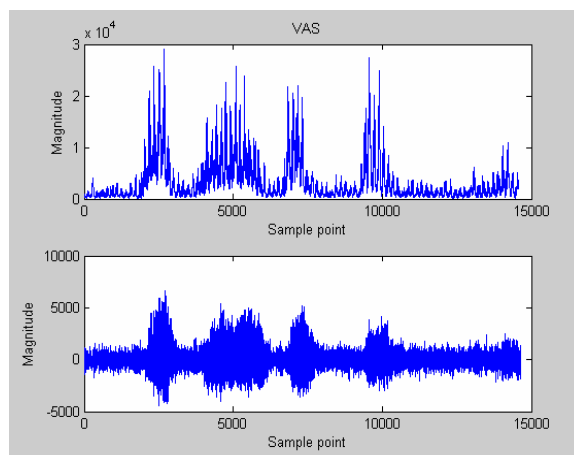


圖 4、VAS 曲線

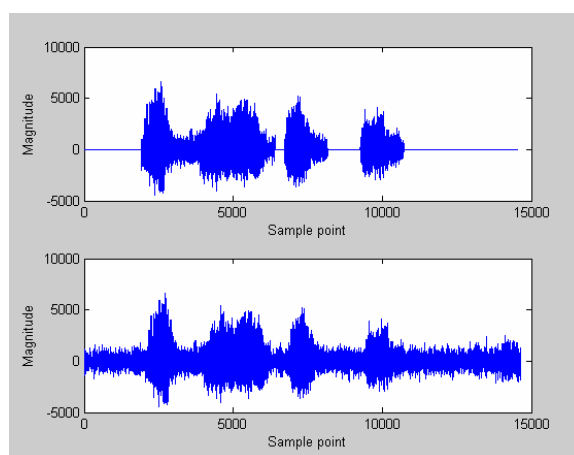


圖 5、本文提出方法實驗結果

表 2、VAD 實驗結果

本文提出的方法				
SNR	Pf	Nf	Nd	Pd
無雜訊	6 %	1%	99%	94%
20dB	7 %	1.1%	98.9%	93%
15dB	8 %	1%	99%	92%
10dB	6.8%	3%	97%	93.2%
5dB	3.2%	9%	91%	96.8%
0dB	1.3%	17.2%	82.8%	98.7%
AMR 語音編碼器中的 VAD 模組				
SNR	Pf	Nf	Nd	Pd
無雜訊	0.34%	2.3%	97.7%	99.66%
20dB	0%	47%	53%	100%
15dB	0%	69%	31%	100%
10dB	0%	97%	3%	100%
5dB	0%	100%	0%	100%
0dB	0%	100%	0%	100%

G.729 語音編碼器中的 VAD 模組				
SNR	P_f	N_f	N_d	P_d
無雜訊	0.5%	10.3%	89.7%	99.5%
20dB	0.1%	79.4%	20.6%	99.9%
15dB	0.1%	78.9%	21.1%	99.9%
10dB	0.23%	79.2%	20.8%	99.77%
5dB	0.1%	84.8%	15.2%	99.9%
0dB	1.4%	81.15%	18.85%	98.6%

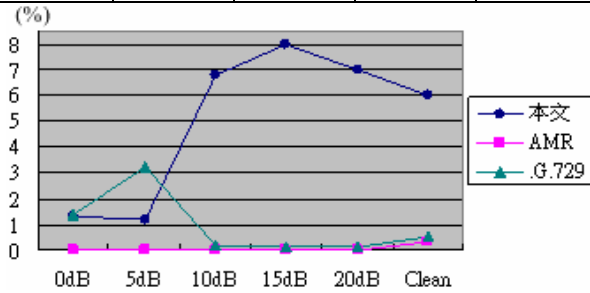


圖 6、VAD 演算法在 P_f 上效能比較

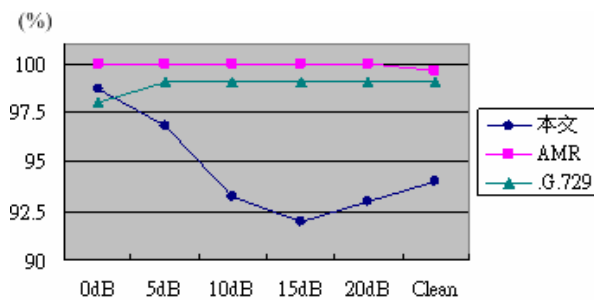


圖 7、VAD 演算法在 P_d 上效能比較

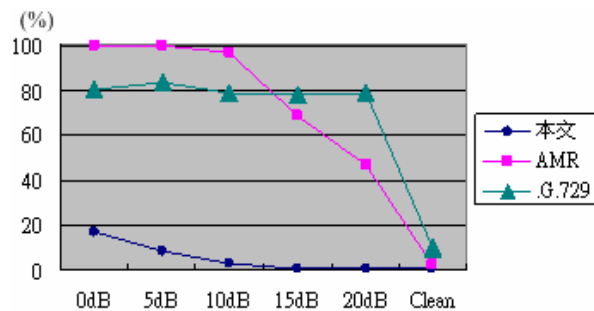


圖 8、VAD 演算法在 N_f 上效能比較

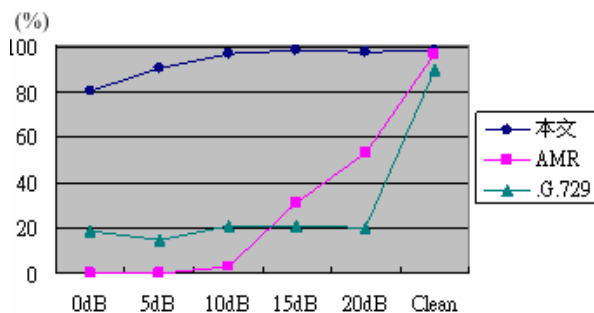


圖 9、VAD 演算法在 N_d 上效能比較

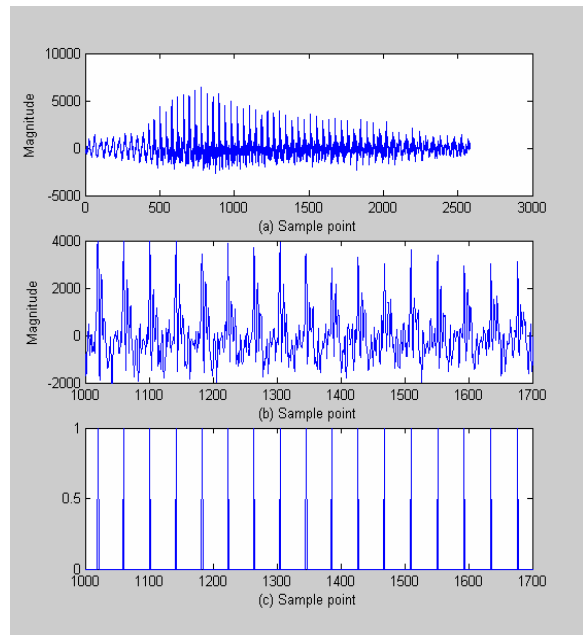


圖 10、Pitch Detection 實驗結果

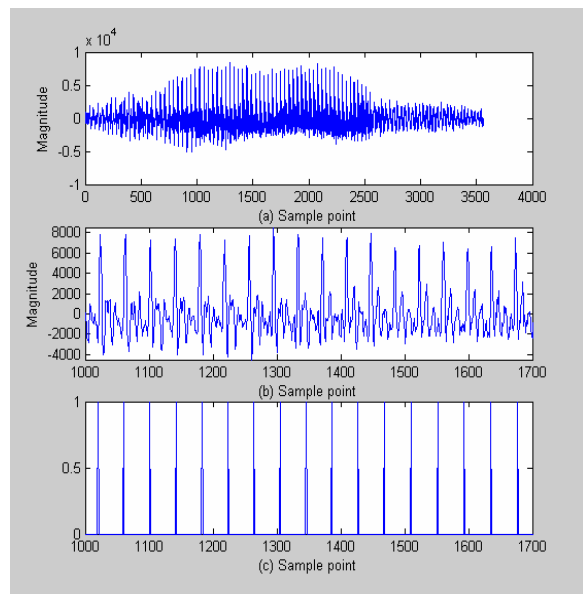


圖 11、Pitch Detection 實驗結果

表 3、Pitch Detection 實驗結果

	本文(男聲)	AMR(男聲)	
SNR	平均校正次數	平均校正次數	平均誤差
無雜訊	412	489	77
20dB	413	491	78
15dB	403	483	80
10dB	418	495	77
5dB	440	523	83
0dB	514	600	86

	本文(女聲)	AMR(女聲)	
SNR	平均校正次數	平均校正次數	平均誤差
無雜訊	459	595	136
20dB	472	606	134
15dB	461	600	139
10dB	479	607	128
5dB	502	630	128
0dB	571	694	123

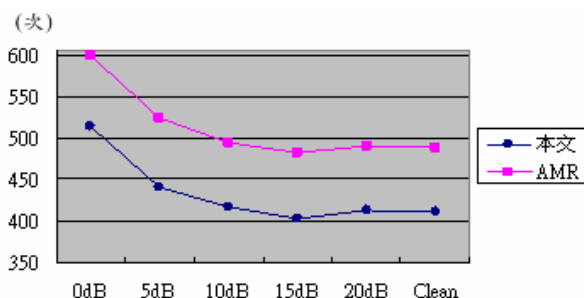


圖12、平均校正次數(男聲)比較

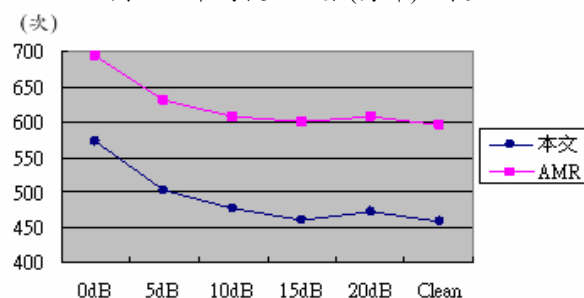


圖13、平均校正次數(女聲)比較

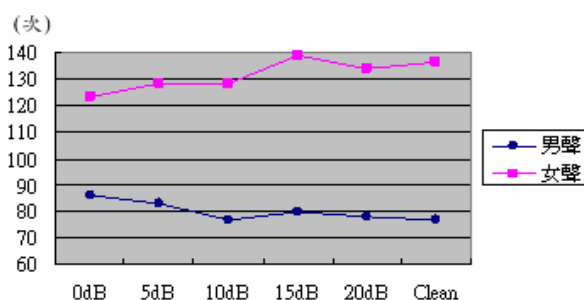


圖14、平均誤差

5. 結論

本文提出的演算法確實可以改善 AMR 中的 VAD 與 Pitch Detection 模組，從實驗結果可以得知，本論文提出的方法在各種語音環境與訊號雜訊下都能夠得到不錯的效能，並且 VAD 不需預設門檻值而是能隨著各種語音環境不同計算出適應性 VAD 門檻值，從實驗可以得知本文提出的 VAD 方

法辨識率比 AMR 預設的還高，而 Pitch Detection 經由實驗可以看出在閉環路修正次數比 AMR 預設的 Pitch Detection 修正還少，證明本文提出的方法找尋 Pitch 週期比 AMR 預設的正確率高。

參考文獻

- [1] E. Ekudden, R. Hagen, I. Johansson, and J. Svedberg, "The adaptive multi-rate speech codec," *IEEE Workshop Speech Coding Proceedings*, pp. 117-119. 1999.
- [2] 陳景川，第三代無線通訊語音編碼之模擬與實現，國立交通大學電子工程研究所，碩士論文。
- [3] Bruno Bessette, Redwan Salami, Roch Lefebvre, Milan Jelínek, Jani Rotola-Pukkila, Janne Vainio, Hannu Mikkola, and Kari Järvinen: "The Adaptive Multirate Wideband Speech Codec (AMR-WB)," *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, VOL. 10, NO. 8, NOVEMBER 2002.
- [4] L. R. Rabiner and M. R. Sambur, "Voiced-unvoiced-silence detection using the Itakura LPC distance measure," in Proc. ICASSP'77, May 1977, pp. 323-326.
- [5] J. C. Junqua, B. Reaves, and B. Mak, "A study of endpoint detection algorithms in adverse conditions: Incidence on a DTW and HMM recognize," in Proc. Eurospeech'91, 1991, pp. 1371-1374.
- [6] J. A. Haigh and J. S. Mason, "Robust voice activity detection using cepstral features," in Proc. *IEEE TENCON*, 1993, pp. 321-324.
- [7] ITU-T Rec. G.729, Annex B, A silence compression scheme for G.729 optimized for terminals conforming to ITU-T V. 70.
- [8] Shi-Huang Chen and Jhing-Fa Wang, "Speech Enhancement Using Perceptual Wavelet Packet Decomposition and Teager Energy Operator"
- [9] Shi-Huang Chen and Jhing-Fa Wang, "A Wavelet-based Voice Activity Detection Algorithm in Noisy Enhancements"
- [10] L. Rabiner and B. H. Juang, *Fundamental of speech recognition*, Upper Saddle River, NJ: Prentice-Hall, 1993.
- [11] J. F. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in Proc. ICASSP'90, 1990, pp. 381-384.
- [12] F. Jabloun, A. E. Cetin, and E. Erzin, "Teager energy based feature parameters for speech recognition in car noise," *IEEE Signal Processing Lett.*, vol. 6, pp.259-261, 1999.
- [13] 吳信德、陳嘉祥、陳璽煌，"以適應性加權門檻值為基礎 VAD 演算法之改良與實作"，2005 數位生活與網際網路科技研討會 2005/2005 年 6 月。
- [14] CHEN, S.H., and WANG, J.F.: "A pyramid-structured wavelet algorithm for detecting pitch period of speech signal,"

Proceedings of 1998 International Computer Symposium (ICS '98), December 1998. pp. 50-56.

- [15] CHEN, S.H., and WANG, J.F.: "Extraction of pitch information in noisy speech using wavelet transform with aliasing compensation," Proceedings of 2001 IEEE international conference on Acoustics, speech, and signal processing (ICASSP 2001), vol. 1, pp. 89-92.
- [16] Shi-Huang Chen, Jhing-Fa Wang, "A Noise-Robust Pitch Detection Method Using Wavelet Transform with Aliasing Compensation," IEE Proc. - Vision, Image and Signal Processing, Vol. 149, No. 6, pp. 327-334, Dec. 2002.