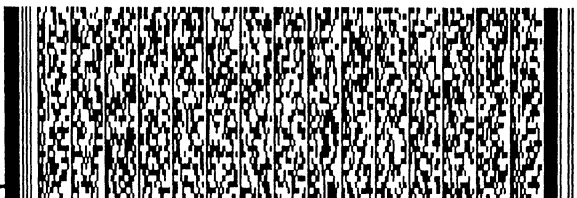


公告本

申請日期：9/12/13	IPC分類 G10L 21/00
申請案號：91136087	

(以上各欄由本局填註) **發明專利說明書** 584835

一、發明名稱	中文	傳送和包裝訊號的數位編碼方法與結構
	英文	METHOD AND ARCHITECTURE OF DIGITAL CODING FOR TRANSMITTING AND PACKING AUDIO SIGNALS
二、發明人 (共2人)	姓名 (中文)	1. 劉啟民 2. 李文傑
	姓名 (英文)	1. Liu, Chi-Min 2. Lee, Wen-Chieh
	國籍 (中英文)	1. 中華民國 TW 2. 中華民國 TW
	住居所 (中文)	1. 新竹縣竹東鎮中興路二段378巷十一號四樓 2. 桃園縣慈文里13鄰慈文路230巷19號
	住居所 (英文)	1. 2.
三、申請人 (共1人)	名稱或姓名 (中文)	1. 國立交通大學
	名稱或姓名 (英文)	1. National Chiao Tung University
	國籍 (中英文)	1. 中華民國 TW
	住居所 (營業所) (中文)	1. 新竹市大學路1001號 (本地址與前向貴局申請者不同)
	住居所 (營業所) (英文)	1.
	代表人 (中文)	1. 張俊彥
	代表人 (英文)	1. Chang, C. Y.



## 一、本案已向

國家(地區)申請專利

申請日期

案號

主張專利法第二十四條第一項優先權

二、主張專利法第二十五條之一第一項優先權：

申請案號：

日期：

三、主張本案係符合專利法第二十條第一項第一款但書或第二款但書規定之期間

日期：

四、有關微生物已寄存於國外：

寄存國家：

寄存機構：

寄存日期：

寄存號碼：

有關微生物已寄存於國內(本局所指定之寄存機構)：

寄存機構：

寄存日期：

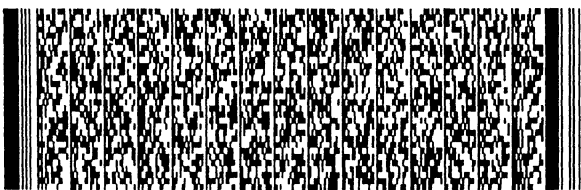
寄存號碼：

熟習該項技術者易於獲得,不須寄存。

## 四、中文發明摘要 (發明名稱：傳送和包裝訊號的數位編碼方法與結構)

一種數位編碼的方法，將輸入音訊轉換成頻率樣本序列，以代表該音訊的一種頻譜合成；並且根據一種位元分配程序，將此頻率樣本序列量化成量化值。藉由一個遮罩門檻值，此位元分配程序利用一個參數預估器來預估量化參數。此量化值被編碼成包含複數個位元的已編碼資料。如果此已編碼資料的位元數超過此已編碼資料的一個預定的可用位元數，則一個反覆比率控制迴圈會調整量化參數和量化間隔大小。在量化頻率樣本序列之前，此方法也會根據一個砍除頻率，將輸入音訊的高頻率分量砍除，此砍除頻率是由反覆比率控制迴圈來決定的。

陸、英文發明摘要 (發明名稱：METHOD AND ARCHITECTURE OF DIGITAL CODING FOR TRANSMITTING AND PACKING AUDIO SIGNALS)

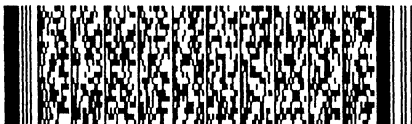


四、中文發明摘要 (發明名稱：傳送和包裝訊號的數位編碼方法與結構)

伍、(一)、本案代表圖為：第\_\_3a\_\_\_\_\_圖

(二)、本案代表圖之元件代表符號簡單說明：

陸、英文發明摘要 (發明名稱：METHOD AND ARCHITECTURE OF DIGITAL CODING FOR TRANSMITTING AND PACKING AUDIO SIGNALS)



## 五、發明說明 (1)

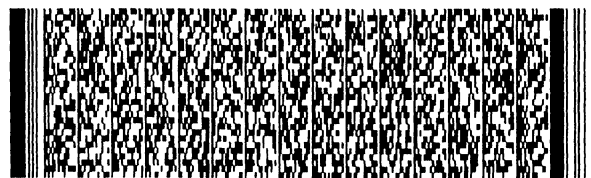
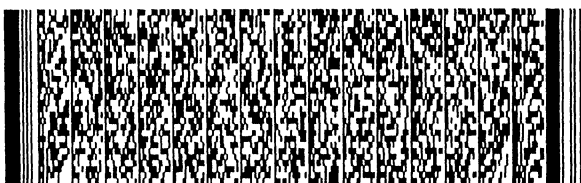
## 發明所屬之技術領域

本發明係有關於一種用於傳送和包裝(packaging)訊號的數位編碼(digital coding)方法與結構，特別是關於音訊(audio signal)編碼中的位元分配(bit allocation)。

## 先前技術

知覺(perceptual)語音編碼如MPEG 1-3 層，進階語音編號，或時間/頻率(time/frequency, T/F)編碼等已廣泛應用在電子電信消費品，電信電訊及播音技術上。在這些知覺語音編碼器中，位元分配是引導高複雜度和決定編碼品質的關鍵模組中一項主要的工作。

圖1為一種知覺音訊編碼中編碼過程(process)的方塊示意圖，時間/頻率轉換器(T/F mapper)101以一個視窗接一個視窗(window-by-window)為基礎將音訊 $S(n)$ 轉換成由時間範圍至頻率範圍的頻段(frequency segments)  $S(m, f)$ 。多種編碼器(coders)103也被用在編碼過程中，以達成高壓縮率(high compression ratios)。輸出值  $X(m, f)$  為以視窗片段指標(window segment index)  $m$  和頻率指標(frequency segment)  $f$  編碼後的序列(sequence)，以頻率範圍為定義域。量化器(quantizer)105將  $X(m, f)$  量

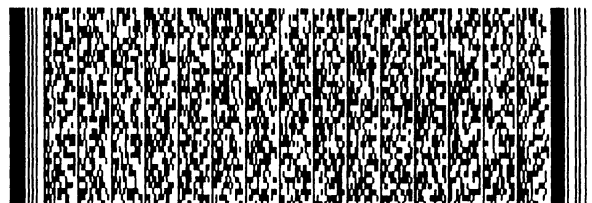
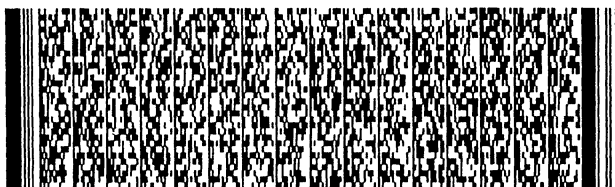


## 五、發明說明 (2)

化成有限個級別(levels)，以 $X'(m, f)$ 表示。並以量化雜音(quantization noise)所產生的相關損害(impairments)減至最小為目的。這些量化級別是由量化參數(quantization parameters)來控制。

一般的音訊壓縮(compression)是將頻率線歸類成組，稱之為量化波段(quantization bands)。一個量化波段所組成的頻率線的數量是根據重要頻波段(critical bands)以及用來傳送量化參數所能提供的位元來決定。可變長度編碼器(variable length coding, VLC)107代表經由可變長度編碼的量化後的序列 $X'(m, f)$ ，其以轉換後的信號的統計上的發生機率來考量。聚集單元(packing unit)109將最終轉碼結果集成一個由特定音訊規範(specified audio protocol)所定義的序列。聽覺分析模式(psychoacoustic model)111將訊號做分析並由分析結果提供訊號-對-遮罩率(signal-to-masking ratio, SMR)給來自訊號分析結果的量化波段。位元分配器(bit-allocator)113參考聽覺分析模式111所提供的遮罩門檻值(masking thresholds)和預定可用位元(available bit budget)115來決定量化參數。

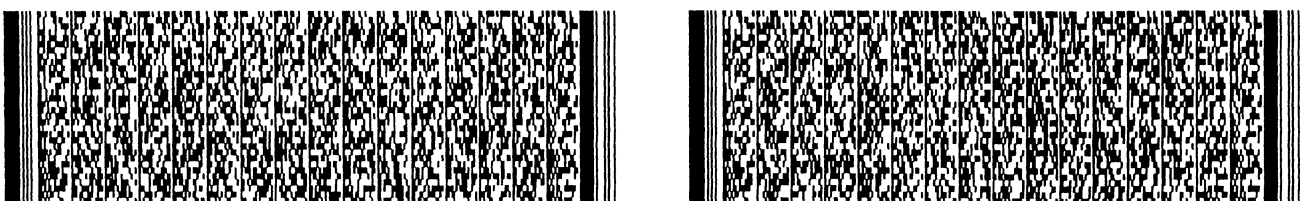
非均勻量化器(non-uniform quantizer)在位元分配器的控制下量化光譜線(spectral lines)，此位元分配器考量最終的音訊品質和需要的位元，來決定量化方式。因



## 五、發明說明 (3)

此，音訊品質和位元數目的控制為位元分配器的基本要件。美國專利文獻5,579,430中揭露了一種有關以頻率定義的最佳編碼(optimum coding in the frequency domain, OCF)過程的數位轉碼程序。它提供一種可媲美CD(compact-disc)品質的音樂轉碼，其可行的資料率(data rate)約為2位元/ATW，以及在良好的FM-無線電廣播品質下1.5位元/ATW的資料率。另一美國專利文獻5,924,060則揭露一種關於聽覺訊號(acoustical signals)傳送及/或儲存的數位編碼程序，其利用一個4和6的因數來減低資料率，而無實質降低音樂訊號的品質。

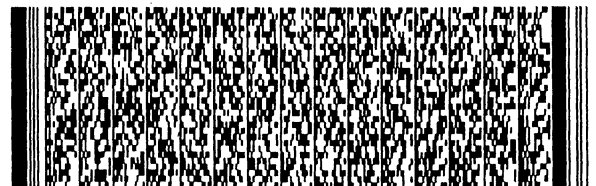
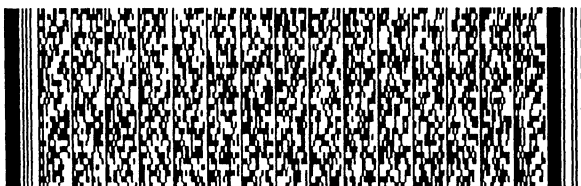
關於MPEG第1和第2層，均勻量化器(uniform quantizer)是用來控制品質和位元要件。因此位元分配器簡單地將可用位元總數分配給次頻波段訊號(sub-band signals)作為量化用，以使量化雜音能聽度(audibility of the quantization noise)最小化。在如MPEG第3層、MPEG-2 AAC和MPEG4 T/F編碼的編碼器中，控制品質和位元率是困難的，主要原因在於它們皆使用非均勻量化器，其量化雜音會隨輸入值而改變。換言之，依據知覺可允許的雜音，它無法藉由指定量化參數來控制品質。用於MPEG第3層和MPEG-2 AAC的可變長度編碼將可變位元長度指定給不同的值，意即被消耗的位元應該是從量化結果中取得，而非從量化器參數本身取得。因此，位元分配是使轉碼器高複雜性的主要工作之一。



## 五、發明說明 (4)

上述缺點造成在估計量化參數時產生問題。一種兩層巢狀(two-nested)迴圈反覆方法(loop iterative method)，稱為OCF，被提出來解決這個問題。如圖2所示，它經由兩種反覆迴圈，即比率控制迴圈(rate-controlling loop)和品質控制迴圈(quality-controlling loop)，來估計量化參數。比率控制迴圈反覆地調整參數值，以符合由量化過程頻譜線的Huffman編碼得到的有限的位元。品質控制迴圈則反覆地調整參數值，以符合反量化過程(inverse quantization)所須評估的量化雜音的知覺標準(perceptual criterion)。

對於一個含有F條頻譜線之畫面(frame)的方法，其複雜度可被描述為 $O(F \cdot R \cdot \eta + F \cdot Q \cdot \gamma)$ ，其中Q和R分別為重覆品質控制迴圈及比率控制迴圈的次數，而 $\eta$ 和 $\gamma$ 又分別為在比率控制迴圈及品質控制迴圈下處理頻譜線的計算複雜度(computation complexity)。比率控制迴圈複雜度 $\eta$ 係來自量化及頻譜線VLC編碼，而品質控制迴圈複雜度 $\gamma$ 則來自解量化(dequantization)和雜音量測(noise measure)。 $\eta$ 和 $\gamma$ 兩者皆為高複雜度。並且，重覆次數Q和R係依量化參數的初始值和調整方法而定，此複雜度甚至大於圖1中混合轉換(hybrid transform)和聽覺分析模式的總複雜度。



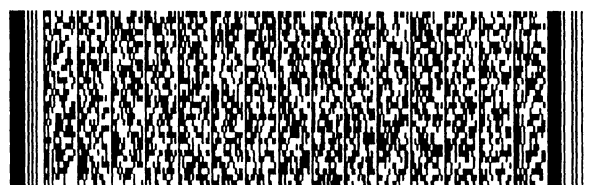


## 五、發明說明 (5)

在品質控制迴圈中指定位元給量化波段決定了語音編碼的品質。指定位元有兩種處理方法。一種方法是只將位元指定給每一反覆迴圈中擁有最差的雜音-對-遮罩率 (noise-to-masking ratio) 的波段。此方法使品質控制迴圈的反覆數目很大，意即相當高的複雜度。另一種方法是在每一反覆迴圈中將位元指定給雜音對遮罩率大於1的所有波段，直到所有可用位元皆被使用殆盡為止。此方法的複雜度較第一種方法的複雜度低許多。然而，令人所關心的是此方法的品質是否讓人滿意。

第一種方法可使雜音形變，以使遮罩門檻值與雜音門檻值相等，此標準已廣被接受。第二種方法曾被使用在ISO的樣本編碼，通常使品質實質上變好。兩種巢狀迴圈法的問題在於它可能無法導致收斂的情況。因為在兩迴圈內控制品質和位元消耗有兩種不同法則，因此可能導致無止盡的迴圈，一般稱之為"停滯"問題 (deadlock problem)。一般用來處理停滯問題的方法是將反覆次數的最大值設一個極限，並且使用某種啟發性的參數調整方法來關照品質和迴圈數。然而，在這些方法中，仍然無法使品質獲得保證。

發明內容

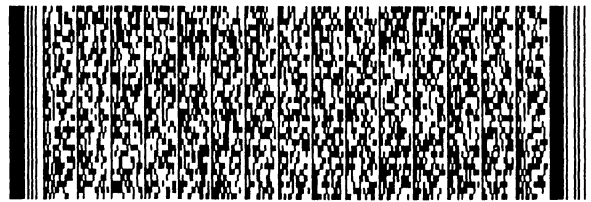
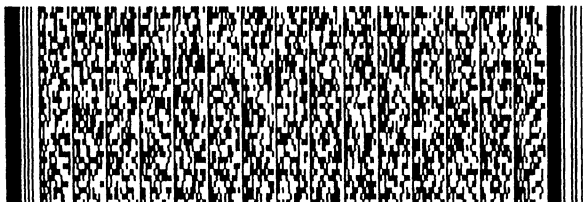


## 五、發明說明 (6)

本發明克服上述傳統數位編碼過程的缺點。其主要目的是提供一種具有高品質及較少複雜度運算之用來傳達和聚集語音訊號的數位編碼方法。

根據本發明，輸入音訊首先被轉換成頻率樣本序列(sequence of frequency samples)，以代表此音訊的頻譜合成(spectral composition)。此頻率樣本序列根據一種位元分配程序和一個參數預估器(parameter predictor)而被量化。此參數預估器直接依據遮罩門檻值來預估量化參數。這些量化值以一可變長度編碼方式來編碼，或是直接集成一個特定的語音規格。如果編碼後的資料的全長超過可用位元的數目，則將參數調整並且增加量化間隔的大小(quantization step size)。重複這個過程直到可用位元的數目大於此編碼所需要的位元數目。最後，將最終已編碼的序列聚集成一個由特定語音規格所定義的序列。

本發明的方法以MPEG第三層的非均勻量化器來詳細說明衍生步驟，且檢查此知覺編碼方法的複雜度和語音品質。依此，本發明利用分段的(segmental)雜音-對-遮罩率說明此衍生步驟，並且提供一種封閉式的方程式(closed-form equation)來表示位元/量化間格大小(bit/step size)和量化語音之間的關係。本方法不限於MPEG第三層；也可應用到如大部份的MPEG AAC (進階語音編碼)



### 五、發明說明 (7)

的知覺編碼器。由於本發明所提供的新的位元分配標準也可以應用到具有均勻量化器的編碼器，如MPEG的第一層和第二層。

本發明的另一目的是提供實現此數位編碼程序的結構。此結構含有一轉換器、一量化器、一可變長度編碼器、一參數預估器、一包裝單元、一調整器(adjustor)，以及一個可被訊號處理器實現以完成本發明之方法的比較器(comparator)。

根據本發明，對於低位元率編碼程序量化參數直接從品質條件來評估。因為藉由利用比率控制迴圈，在非均等的頻率線裡的量化頻寬(quantization bandwidth)和所需的位元數被考量，並且可被調整。而對於可變位元率的編碼，則可完全除去比率控制迴圈。

茲配合下列圖示、實施例之詳細說明及專利申請範圍，將上述及本發明之其他目的與優點詳述於后。

### 實施方式

圖3a說明本發明之音訊編碼方法的程序。參照圖3a，輸入音訊首先被轉換成一頻率樣本序列，以代表此音訊的

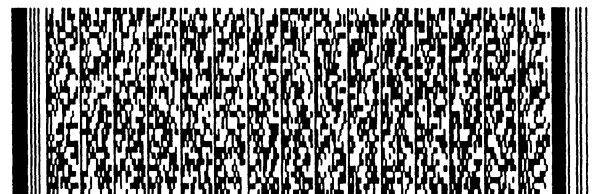
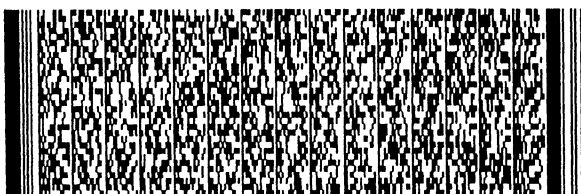


## 五、發明說明 (8)

頻譜合成。接著，此頻率樣本序列根據一種位元分配程序，而被量化以獲得具有一個較低精確度的符號。一個參數預估器直接依據遮罩門檻值來預估量化參數，此遮罩門檻值為人類聽覺系統所能聽見的雜音限度。然後，預估一個壓縮系統之決定訊號級別解析度的參數。

這些已量化符號的編碼是以一個可變長度編碼器來編碼。下一個步驟則是檢查一個預定的可用位元數是否足夠讓已編碼資料使用。如果編碼後的資料的全長超過可用位元的數目，則將參數調整並且增加量化間隔的大小。重複這個過程直到可用位元的數目大於此編碼所需要的位元數目。最後，將最終已編碼的序列包裝成一個由特定音訊規格所定義的序列。

對於低位元率音訊的編碼，在評估參數預估器中的量化參數之前，高頻率可被砍除(cut-off)。圖3b說明此低位元率音訊編碼程序的步驟。如圖3b所示，當低位元率語音編碼所需要的位元數目超過可用位元的數目時，調整砍除頻率(cut-off frequency)然後傳送，以使高頻率的分量在預估量化參數之前就被砍除。如有需要，則量化間隔大小也將被調整。對於可變位元率音訊的編碼，可用位元可根據所需要的品質而被調整。在此情況下，反覆比率控制迴圈的步驟則可以完全被移除。圖3c說明此可變位元率音訊編碼程序的步驟，其中反覆比率控制迴圈的步驟從圖

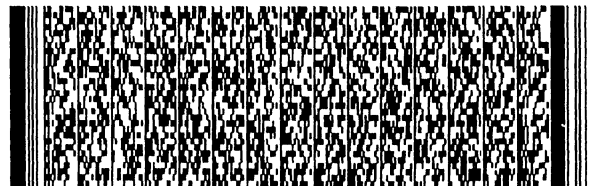
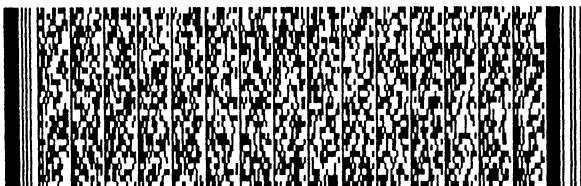


## 五、發明說明 (9)

3a 中被移除。

本發明之圖3a到圖3c的編碼程序可用訊號處理器 (signal processors) 來實現。此實現的詳細架構揭露如后。根據圖3a，圖4a所示的實現架構包含有一轉換器401，將輸入音訊接收並轉換成一頻率樣本序列，以代表該音訊的一種頻譜合成。一個量化器402，根據一個位元分配程序，將此頻率樣本序列量化成有限個級別(level)。一個參數預估器405，直接藉由一個遮罩門檻值來估計量化參數，然後一個最佳的編碼器403將已量化的級別編碼。當預定預定可用位元數不足夠讓已編碼資料使用時，一個調整器407調整這些量化參數。一個比較器408，比較一個預定的可用位元數和已編碼資料所需的長度，以檢查此可用的位元數是否足夠讓已編碼資料使用。一個包裝單元409，將最終已編碼的序列包裝成一個由特定音訊規格所定義的序列。

圖4b和圖4c分別為實現圖3b和3c的架構。參考圖4b，一個調整器413用來調整砍除的頻率，並在低位元率音訊編碼的情形下，將此砍除的頻率傳送到高頻率砍除單元411。調整器413也可以調整用在量化器402裡的量化間隔大小。高頻率砍除單元(high-frequency cut-off unit) 411是被加在轉換器401和量化器402之間，用來接收調整後的砍除頻率，並將它傳送給參數預估器405。在可變位



## 五、發明說明 (10)

元率編碼的情形下，與反覆比率控制迴圈步驟有關的元件均被移除，如圖4c所示。

在本發明中，產生一個以常數遮罩-對-雜音率  $\rho$  為基礎的決定公式(deterministic formula)，來計算在位元分配程序中參數預估器的量化參數。此決定方程式提供一個雜音預估器的封閉式公式給非均勻量化器。本發明的方法以MPEG第三層來詳細說明衍生步驟和作為實施例。對於一個MPEG ACC量化器，也可以使用類似的程序。

本發明的位元分配藉由單一步驟的預測，就符合了對於每一個子波段(sub-band)的位元率和雜音成形(noise shaping)的需求。而直接藉由一個遮罩門檻值，最佳的全域因子(global factor)和每一個子波段的比率調整因子(scaling factor)就被預估。此全域因子控制了被使用殆盡之位元的總數，而此比率調整因子則控制了該相關波段相對於其他波段的量化雜音。以下章節首先說明此位元分配的要件，再詳加說明雜音預估器(noise predictor)的衍生步驟，以及在零波段(zero band)和負值雜音遮罩率(negative noise-to-masking ratio, NMR)條件下的一個比率調整因子的上下限(bounds)。

位元分配條件

## 五、發明說明 (11)

首先，考慮分段的NMR的最小值，並表示為：

$$R(i) = \arg \underset{R(i)}{\text{Min}} \sum_i \left\{ \left( \frac{\sigma_{N(i)}^2}{\sigma_{M(i)}^2} \right) \right\}, \quad (1)$$

其中  $\sigma_{N(i)}^2$  和  $\sigma_{M(i)}^2$  分別為與關鍵波段  $i$  相關的雜音能量 (noise energy) 和遮罩能量 (masking energy)。  $R(i)$  為將 NMR 最小化的位元率。在一個每樣本有  $R(i)$  個位元的 PCM 編碼器中，量化誤差變異數 (quantization error variance) 表示如下：

$$N(i) = \rho 2^{-2R(i)} \sigma_{x(i)}^2 \quad (2)$$

因此，此最小值

$$\arg \underset{R(i)}{\text{Min}} \sum_i \left\{ \left( \frac{\rho 2^{-2R(i)} \sigma_{x(i)}^2}{\sigma_{M(i)}^2} \right) \right\} \quad (3)$$

必須受限於此總位元率，亦即，

$$\sum_i \{R(i)B(i)\} = R. \quad (4)$$

根據拉氏乘數 (Lagrange multipliers) 的方法，其  $\lambda$  解必須滿足下列方程式：



## 五、發明說明 (12)

$$\text{對於所有 } j, \quad \frac{\partial}{\partial R(j)} \left\{ \left( \sum_i \{R(i)B(i)\} - R \right) + \lambda \sum_i \left\{ \left( \frac{\rho 2^{-2R(i)} \sigma_{x(i)}^2}{\sigma_{M(i)}^2} \right) \right\} \right\} = 0,$$

$$\text{則 } \lambda = \frac{B(j)}{(2 \log 2) \left( \frac{\rho 2^{-2(R(j))} \sigma_{x(j)}^2}{\sigma_{M(j)}^2} \right)} = \frac{B(j)}{2 \log 2 \left( \frac{\sigma_{N(j)}^2}{\sigma_{M(j)}^2} \right)}, \quad \text{對於所有 } j \quad (5)$$

所以，必須找出  $R(j)$  使得雜音遮罩率正比於此  $B(j)$ 。亦即，

$$\sigma_{N(j)}^2 = \kappa \sigma_{M(j)}^2 B(j), \quad \text{對於所有 } j \quad (6)$$

雜音的標準水平(noise level)必須正比於遮罩門檻值乘以一個頻寬，以便具有最佳分段的NMR。

其次，考量在該量化波段的遮罩門檻值和關鍵頻寬以選取量化波段的雜音標準水平。換句話說，將找出  $\sigma_{N(q)}^2$  來取代  $\sigma_{N(j)}^2$  以使分段的NMR最小化。

$$\sigma_{N(q)}^2 = \kappa \sigma_{M(q)}^2 B(q) \quad (7)$$

其中， $q$ 是量化波段的指標。此問題就是找出  $B(q)$  來最佳逼近使分段的NMR最小化的能量，亦即

$$\hat{B}(q) = \arg \text{Min}_{B(q)} \sum_{j \in q} \left| \sigma_{N(q)}^2 - \sigma_{N(j)}^2 \right| \quad (8)$$





## 五、發明說明 (13)

假設在量化波段裡的關鍵波段的遮罩能量是均勻的，則經計算後的選擇為

$$\hat{B}(q) = \underset{j \in q}{\text{Average}}(B(j)) \quad (9)$$

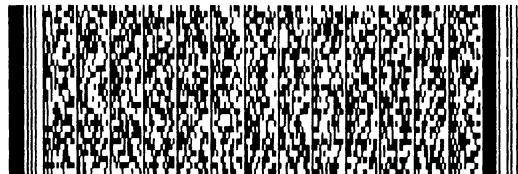
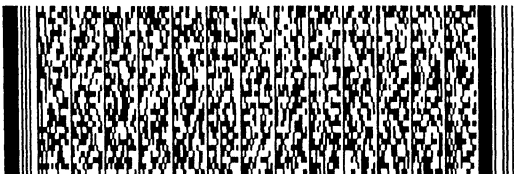
第三，為了避免位元被分配給具有高於雜音水平標準的遮罩水平標準的波段，必須修改將分段的NMR最小化的條件，使得具有負值NMR的波段被歸整為1。意即，每一波段的量化雜音必須有一個下限。另一方面，高於遮罩門檻值的雜音會導致一種現象，此現象即相關波段將被歸整為零，稱之為零波段。這些零波段是相當明顯的。所以，這些量化級別也必須被限制為不可大於訊號能量。

總而言之，在零波段和負值NMR的條件下，指定此位元分配時，雜音必須與遮罩門檻值乘以頻寬的值相等。

雜音預估器

本發明以MPEG第三層的量化器來詳細說明雜音分配器的衍生步驟。從MPEG第三層的標準，此第三層之非均勻量化器的簡化公式為

$$is_i = \text{int} \left( \frac{xr_i^{\frac{3}{4}}}{\Delta_q} \right), \quad (10)$$



## 五、發明說明 (14)

其中，量化間隔大小為

$$\Delta_{sfb} = 2^{\frac{3}{4}(gain_{gr} - scale_{sfb})} \quad (11)$$

從MPEG的標準，非均勻量化器的公式也可以表示為

$$is_i = \text{int} \left( xr_i 2^{\frac{3}{4}(scale_q - gain_{gr} - 0.0946)} \right) \quad (12)$$

其中，對於每個量化波段q，比率調整因子為

$scale_q = 1/2(1 + scalefac\_scale)(scalefac_q + preflag \cdot pretab_q)$ ； $scalefac\_scale$  為0或1，

$scalefac_q$  介於0到15之間，而預先放大的旗幟

(pre-amplified flag) 為  $preflag_{gr} \cdot pretab_q$ ；對於MPEG第三

層結構的每一個小細節(granule)，全域益(global gain)

為  $gain_{gr} = 1/2(global\_gain_{gr} - 210)$ 。將0.0946忽略，則(12)可被

推導成

$$\begin{aligned} is_i &= \text{int} \left( xr_i 2^{\frac{3}{4}(scale_q - gain_{gr})} \right) \\ &= \text{int} \left( xr_i^{\frac{3}{4}} 2^{\frac{3}{4}(scale_q - gain_{gr})} \right) \\ &= \text{int} \left( \frac{xr_i^{\frac{3}{4}}}{\Delta_q} \right) \end{aligned} \quad (13)$$

其中，間隔大小為  $\Delta_q = 2^{\frac{3}{4}(gain_{gr} - scale_q)}$ 。



## 五、發明說明 (15)

接下來，輸入訊號(input signal)  $x_{r_i}$  和回復訊號(reconstructed signal)  $\tilde{x}_{r_i}$  有下列兩個公式來表示：

$$x_{r_i} = (is_i + \varepsilon_i) \Delta_{sfb}^{\frac{4}{3}} \quad , \quad \text{和} \quad \tilde{x}_{r_i} = (is_i \Delta_{sfb})^{\frac{4}{3}}$$

非均勻量化器  $e_i$  的量化誤差將等於輸入訊號  $x_{r_i}$  和回復訊號  $\tilde{x}_{r_i}$  的差值：

$$\begin{aligned} e_i &= x_{r_i} - \tilde{x}_{r_i} = \left( (is_i + \varepsilon_i) \Delta_{sfb} \right)^{\frac{4}{3}} - \left( is_i \Delta_{sfb} \right)^{\frac{4}{3}} \\ &= (1 + is_i^{-1} \varepsilon_i)^{\frac{4}{3}} is_i^{\frac{4}{3}} \Delta_{sfb}^{\frac{4}{3}} - \left( is_i \Delta_{sfb} \right)^{\frac{4}{3}} \end{aligned} \quad (14)$$

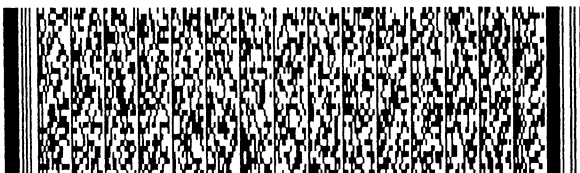
令  $f(\varepsilon_i) = (1 + is_i^{-1} \varepsilon_i)^{\frac{4}{3}}$ 。以泰勒展開式(Tyler expansion)的第一階導數  $f(\varepsilon) \approx 1 + f'(\varepsilon)\varepsilon$  來逼近此差值，而導出

$$e_i = f(\varepsilon_i) is_i^{\frac{4}{3}} \Delta_q^{\frac{4}{3}} - \left( is_i \Delta_q \right)^{\frac{4}{3}} \approx \frac{4}{3} is_i^{\frac{1}{3}} \varepsilon_i \Delta_q^{\frac{4}{3}}$$

假設量化訊號  $is_i$  和非均勻量化器  $\varepsilon_i$  的量化後的誤差相互獨立，則非均勻量化器  $e_i$  的量化誤差的期望值如下：

$$E[e_i^2] \approx \frac{16}{9} \Delta_q^{\frac{8}{3}} E[IS_i^{\frac{2}{3}} \mathcal{E}^2] \approx \frac{16}{9} \Delta_q^{\frac{8}{3}} E[IS_i^{\frac{2}{3}}] E[\mathcal{E}^2] \quad (15)$$

如果量化波段的光譜是均勻的，則線的雜音可以是此量化波段的平均能量；亦即



五、發明說明 (16)

$$E(e_i^2) = E(e_q^2) \quad (16)$$

因為  $E[\varepsilon_i^2] = \frac{1}{12}$ ，(15) 式變成

$$E[e_i^2] \approx \frac{4}{27} \Delta_q^{\frac{8}{3}} E\left[\left(\frac{XR_i^{\frac{3}{4}}}{\Delta_q}\right)^{\frac{2}{3}}\right] = \frac{4}{27} \Delta_q^2 E[|XR_i|^{\frac{1}{2}}] \quad (17)$$

將(7)式帶入(16)式得出

$$E[e_q^2] = \kappa \sigma_{M(q)}^2 B(q) \quad (18)$$

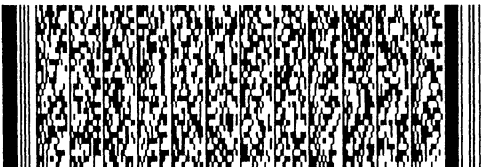
最後，藉由定義  $T_q = \sigma_{M(q)}^2 B(q)$ ，此全域益和比率調整因子之間的差值趨近至

$$gain_{gr} - scale_q \approx \frac{2}{3} \log_2 \frac{27}{4} \kappa \cdot T_q^2 / E[|XR_q|^{0.5}],$$

或

$$gain_{gr} - scale_q = \frac{2}{3} \left( \log_2 \frac{27}{4} + \log_2 \kappa + \log_2 T_q^2 - \log_2 E[|XR_q|^{0.5}] \right) \quad (19)$$

因為比率調整因子  $scale_q$  的值介於0和16之間，並且這些量化波段之比率調整的最小值必須為零，因此全域益為



## 五、發明說明 (17)

$$gain_{gr} = \text{Max}_q \{ gain_{gr} - scale_q \},$$

並且得到所有子波段的比率調整因子。可見，全域益隨著位元率相關常數  $\kappa$  而改變，而每一子波段的比率調整因子則根據此遮罩門檻值和輸入訊號而不同。

比率調整因子的上下限

如前所述，位元分配應受限於非負值NMR和零波段的條件下。對於非負值NMR的情形，雜音級別被設定為遮罩門檻值；亦即  $T_q = \sigma_{M(q)}^2$  和  $\kappa = 1$ 。如此產生相對於全域比率調整之  $Uscale_q$  的上限。

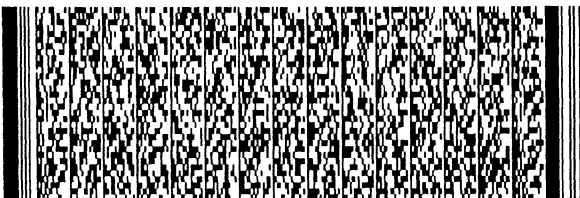
$$gain_{gr} - Uscale_q = \frac{2}{3} \left( \log_2 \frac{27}{4} + \log_2 \sigma_{M(q)}^2 - \log_2 E[|XR_q|^{0.5}] \right) \quad (20)$$

意即，

$$scale_q \leq Uscale_q = gain_{gr} - \frac{2}{3} \left( \log_2 \frac{27}{4} + \log_2 \sigma_{M(q)}^2 - \log_2 E[|XR_q|^{0.5}] \right) \quad (21)$$

此  $gain_{gr}$  將根據可用的位元而被調整。

下限可由零波段的條件下導出。當雜音大於訊號能量時，則會發生零波段的情形；亦即



## 五、發明說明 (18)

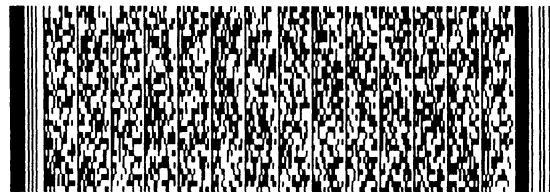
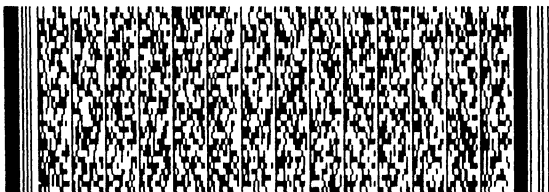
$$\Delta_q^2 = (2^{\frac{3}{4}(gain_{gr} - Dscale_q)})^2 < \left\{ E[|XR_q|^{0.5}] \right\}^4 \quad (22)$$

因此，比率調整的下限會是

$$scale_q \geq Dscale_q = gain_{gr} - \frac{1}{2} \log_2 E[|XR_q|^{0.5}] \quad (23)$$

圖5分別說明本發明和MPEG位元分配程序使用不同試驗材料的平均反覆次數，其中Q為品質控制的平均反覆次數，而R為比率控制的平均反覆次數。如圖5所示，本發明的分配方法移除了品質控制反覆所需要的反覆次數，並且利用一個大於3的因子來減少比率控制的反覆次數。

圖6說明本發明相較於ISO之位元分配方法的客觀的評分紀錄。在此，本發明採用語音品質的知覺預估值(perceptual evaluation of audio quality, PEAQ)系統，此PEAQ系統為ITU-R工作組10/4所推薦的系統。ISO為原始碼。ISO1採用Lame的終止條件而被改善。此實驗以立體款式(stereo mode)和心理分析模式2為基礎。並且，由於MS開關和位元保留器(bit reservoir)與位元分配方法無關，因此實驗中將此兩個機制關掉。客觀差異等級(objective difference grade, ODG)是客觀測量方法的輸出變數。ODG值應理想地介於0和-4之間，其中0對應一個察覺不出的損害，而-4係對應一個被認定是非常惱人的



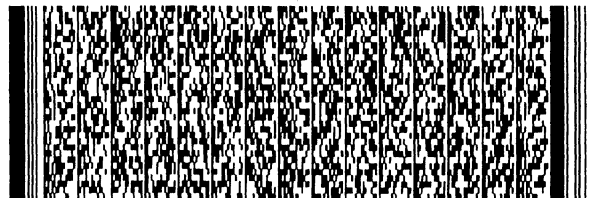
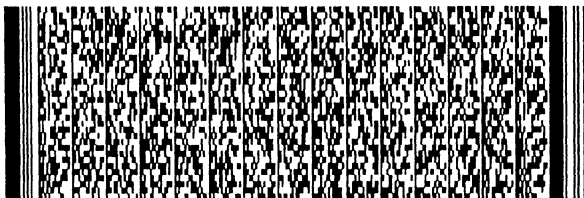
## 五、發明說明 (19)

損害。如圖6所示，本發明之方法的品質較此圖文中所提出的方法品質為好。

PEAQ在本發明中所採用的架構為基本的版本。此基本版本使用以FFT為基礎的耳朵模型(FFT-based ear model)。它使用下列的模型輸出變數：BandwidthRef<sub>B</sub>，BandwidthTest<sub>B</sub>，Total NMR<sub>B</sub>，WinModDiff1<sub>B</sub>，ADB<sub>B</sub>，EHS<sub>B</sub>，AvgModDiff1<sub>B</sub>，AvgModDiff2<sub>B</sub>，RmsNoiseLoud<sub>B</sub>，MFPD<sub>B</sub>以及RelDistFrames<sub>B</sub>。利用一個人造類神經網(artificial neural network)，這11個模型輸出變數被轉換成一個單一品質指數，此人造類神經網的隱藏層(hidden layer)中含有三個節點(node)。

圖7提供一組測試訊號的表單，這些測試訊號被使用於客觀和主觀的試驗中。藉由設定相同的反覆終止條件，例如反覆次數、非增加雜音比率調整因子波段、合適於比率調整因子表等等[網址 <http://www.mp3dev.org/mp3>]，ISO演譯法則可以利用Lame提到的方法來改善(即具有最佳品質的mp3編碼器)。被採用來比較的兩個巢狀迴圈係以Lame所使用的反覆演譯法則為基礎。

唯，以上所述者，僅為本發明之較佳實施例而已，當不能以此限定本發明實施之範圍。即大凡依本發明申請專利範圍所作之均等變化與修飾，皆應仍屬本發明專利涵蓋



五、發明說明 (20)

之範圍內。





圖式簡單說明

圖1係時下音訊編碼之一種編碼程序的方塊示意圖。

圖2係一OCF程序中的位元分配程序。

圖3a說明本發明之音訊編碼程序的流程步驟。

圖3b說明本發明之低位元率音訊編碼程序的流程步驟。

圖3c說明本發明之可變位元率音訊編碼程序的流程步驟。

圖4a說明本發明之圖3a的一種實現架構。

圖4b和圖4c分別說明圖3b和圖3c的實現架構。

圖5分別說明本發明和MPEG位元分配程序使用不同試驗材料時，品質控制和位元率控制的平均反覆次數。

圖6說明本發明相較於ISO之位元分配方法的客觀的評分紀錄。

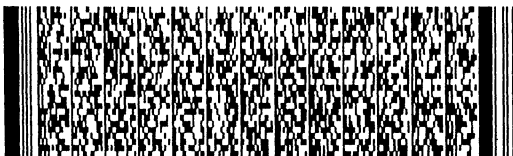
圖7提供一組測試訊號的表單，這些測試訊號被使用於客觀和主觀的試驗中。



## 圖式簡單說明

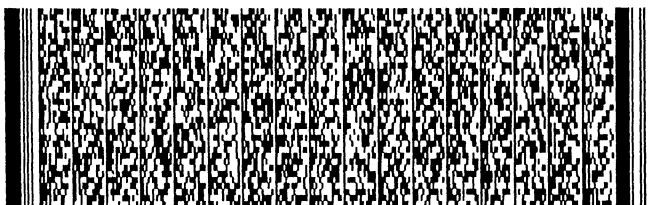
## 圖號說明

101	時間/頻率轉換器	103	其他編碼器
105	量化器	107	可變長度編碼器
109	包裝單元	111	聽覺分析模式
113	位元分配器	115	可用的位元
401	轉換器	402	量化器
403	可變長度編碼器	405	參數預估器
407	調整器	408	比較器
409	包裝單元	411	砍除高頻率
413	調整器		



## 六、申請專利範圍

1. 一種用於傳送和包裝音訊之數位編碼的方法，包含下列步驟：
  - (a) 將輸入音訊轉換成頻率樣本序列，以代表該音訊的一種頻譜合成；
  - (b) 根據一種位元分配程序，將該頻率樣本序列量化成量化值，藉由一個遮罩門檻值，該位元分配程序利用一個參數預估器來估計量化參數；
  - (c) 利用一個符號編碼器將該量化值編碼以形成已編碼的資料，該已編碼的資料包含複數個位元；以及
  - (d) 根據一個特定音訊規格，將該已編碼的資料包裝成一個資料序列。
2. 如申請專利範圍第1項所述之用於傳送和包裝音訊之數位編碼的方法，其中該步驟(b)係由一個均勻量化器或是一個非均勻量化器來執行。
3. 如申請專利範圍第1項所述之用於傳送和包裝音訊之數位編碼的方法，其中該符號編碼器包含一個可變長度編碼器。
4. 如申請專利範圍第1項所述之用於傳送和包裝音訊之數位編碼的方法，其中對於一個量化波段，該位元分配程序中的參數預估器利用一個公式來計算並調整至少



## 六、申請專利範圍

一個對應的全域因子和/或一頻波段比率調整因子，該公式係以一個常數遮罩-對-雜音率為基礎。

5. 如申請專利範圍第4項所述之用於傳送和包裝音訊之數位編碼的方法，其中該步驟(b)之位元分配程序更包含下列步驟：

根據該已編碼資料的一個預定的可用位元數來調整該全域因子；以及

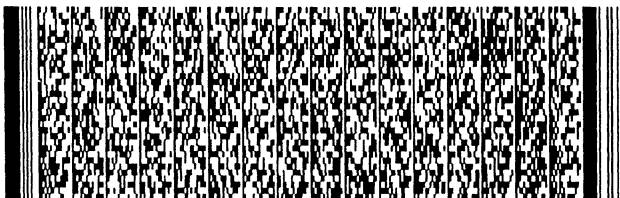
產生與該量化波段全域因子對應的該波段比率調整因子的一個上限和一個下限。

6. 如申請專利範圍第5項所述之用於傳送和包裝音訊之數位編碼的方法，其中該上限是由一個非負值的噪音-對-遮罩率來限制。

7. 如申請專利範圍第5項所述之用於傳送和包裝音訊之數位編碼的方法，其中該下限是由零波段來限制。

8. 如申請專利範圍第4項所述之用於傳送和包裝音訊之數位編碼的方法，其中根據該遮罩門檻值和該輸入的音訊，該波段比率調整因子依每個子波段而不同。

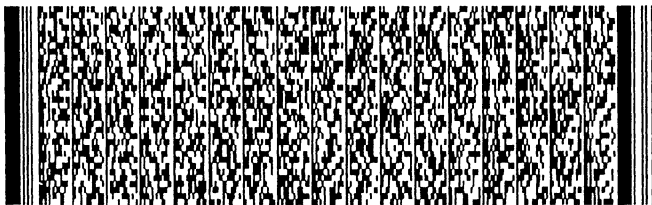
9. 如申請專利範圍第4項所述之用於傳送和包裝音訊之數位編碼的方法，其中該全域因子隨一個位元率相關的



## 六、申請專利範圍

常數而不同。

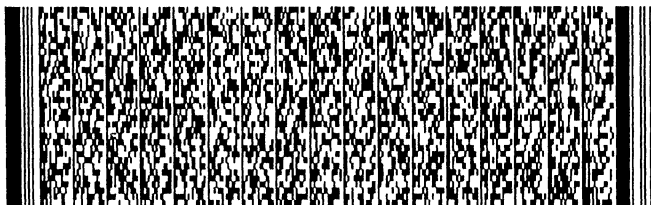
10. 如申請專利範圍第1項所述之用於傳送和包裝音訊之數位編碼的方法，在步驟(d)之前，更備有一個反覆比率控制迴圈，該反覆比率控制迴圈包含下列步驟：
- (c1) 如果包含於該已編碼資料的該位元數不超過該已編碼資料的一個預定的可用位元數，則繼續該步驟(d)，否則繼續步驟(c2)；以及
- (c2) 調整量化參數和一個將用於步驟(b)的量化間隔大小，並且回到步驟(b)。
11. 如申請專利範圍第10項所述之用於傳送和包裝音訊之數位編碼的方法，其中該步驟(b)係經由一均勻量化器或是一個非均勻量化器來執行。
12. 如申請專利範圍第10項所述之用於傳送和包裝音訊包裝音訊之數位編碼的方法，其中如果包含於該已編碼資料的該位元數超過該已編碼資料的一個預定的可用位元數，則至少一個對應的全域因子和一個頻波段比率調整因子會被調整，並且該步驟(c2)中的量化間隔大小也會增加。
13. 如申請專利範圍第1項所述之用於傳送和包裝音訊之數位編碼的方法，其中該符號編碼器包含一個可變長



## 六、申請專利範圍

度編碼器。

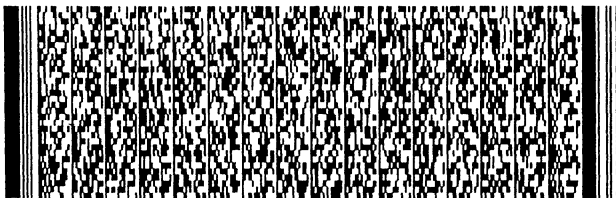
14. 如申請專利範圍第10項所述之用於傳送和包裝音訊之數位編碼的方法，其中該步驟(b)更包含一個步驟：在量化該頻率樣本序列之前，對於低位元率音訊的編碼，砍除高頻率。
15. 如申請專利範圍第14項所述之用於傳送和包裝音訊之數位編碼的方法，其中該反覆比率控制迴圈的步驟(c2)更包括對該砍除高頻率的步驟去調整一個砍除頻率的步驟。
16. 如申請專利範圍第10項所述之用於傳送和包裝音訊之數位編碼的方法，其中對於一個量化波段，該位元分配程序中的參數預估器利用一個公式，來計算並調整至少一個對應的全域因子和/或一頻波段比率調整因子，該公式係以一個常數的遮罩-對-雜音率為基礎。
17. 如申請專利範圍第16項所述之用於傳送和包裝音訊之數位編碼的方法，其中該步驟(b)之位元分配程序更包含下列步驟：  
根據該已編碼資料的一個預定的可用位元數來調整該全域因子；以及  
產生與該量化波段全域因子對應的該波段比率調整因



## 六、申請專利範圍

子的一個上限和一個下限。

18. 如申請專利範圍第17項所述之用於傳送和包裝音訊之數位編碼的方法，其中該上限是由一個非負值的噪音-對-遮罩率來限制。
19. 如申請專利範圍第17項所述之用於傳送和包裝音訊之數位編碼的方法，其中該下限是由零波段來限制。
20. 如申請專利範圍第16項所述之用於傳送和包裝音訊之數位編碼的方法，其中根據該遮罩門檻值和該輸入的音訊，該波段比率調整因子依每個子波段而不同。
21. 如申請專利範圍第16項所述之用於傳送和包裝音訊之數位編碼的方法，其中該全域因子隨一個位元率相關的常數而不同。
22. 一種用於傳送和包裝音訊之數位編碼的結構，該結構包含：
  - 一個轉換器，該轉換器將輸入音訊轉換成一頻率樣本序列，以代表該音訊的一種頻譜合成；
  - 一個參數預估器，該參數預估器藉由一個遮罩門檻值來估計量化參數；
  - 一個量化器，該量化器根據該量化參數，將該頻率樣



## 六、申請專利範圍

本序列量化成量化值；

一個可變長度編碼器，該可變長度編碼器將該量化值編碼成已編碼的資料，其中該已編碼的資料包含複數個位元；以及

一個包裝單元，該包裝單元根據一個特定音訊規格，將該已編碼的資料包裝成一個資料序列。

23. 如申請專利範圍第22項所述之用於傳送和包裝音訊之數位編碼的結構，該結構更包含：

一個比較器，該比較器比較包含於該已編碼資料的該位元數和該已編碼資料的一個預定的可用位元數；以及

一個調整器，該調整器當包含於該已編碼資料的該位元數超過該已編碼資料的該預定的可用位元數時，用來調整該量化參數。

24. 如申請專利範圍第23項所述之用於傳送和包裝音訊之數位編碼的結構，該結構更包含一個高頻率砍除單元，該高頻率砍除單元連接在該轉換器和該量化器之間，並且備有一個輸入口，用來接收來自該調整器的一個砍除頻率。





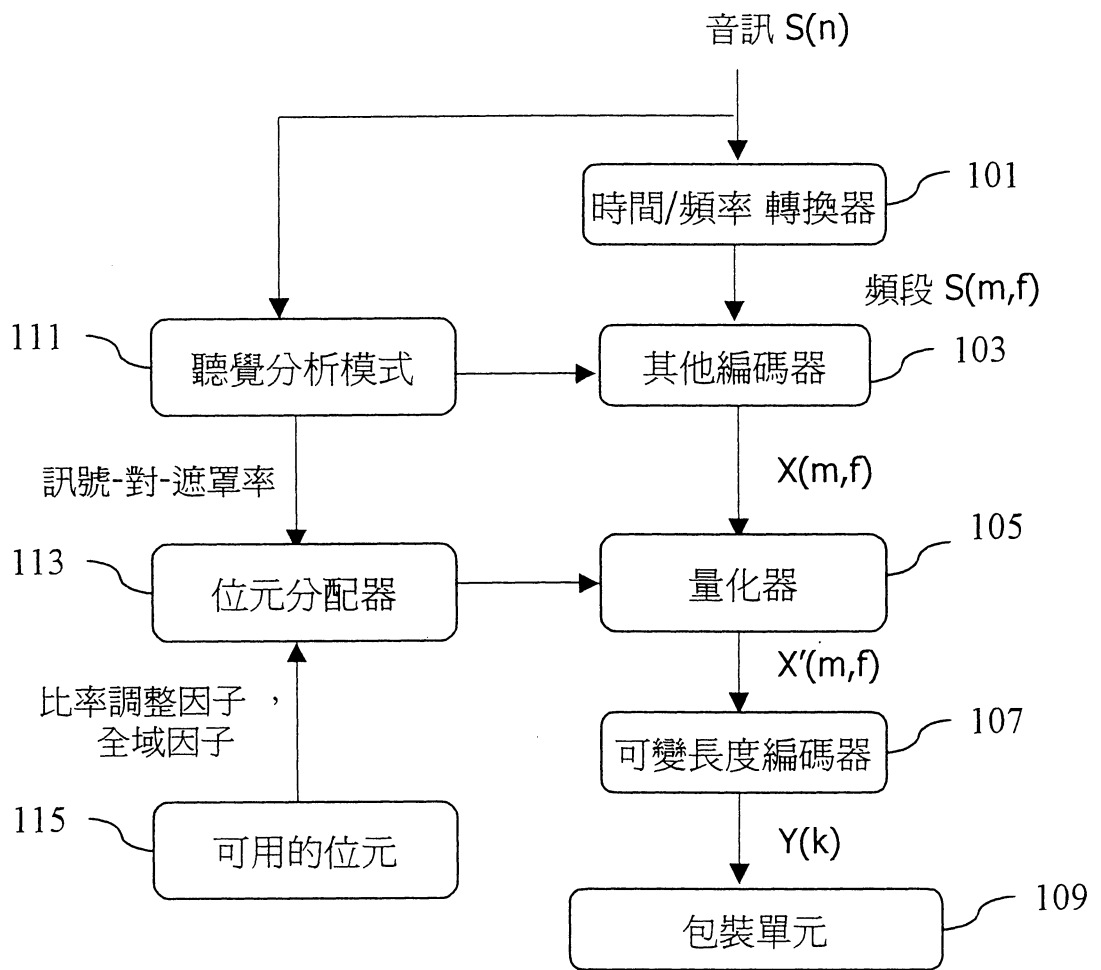


圖 1

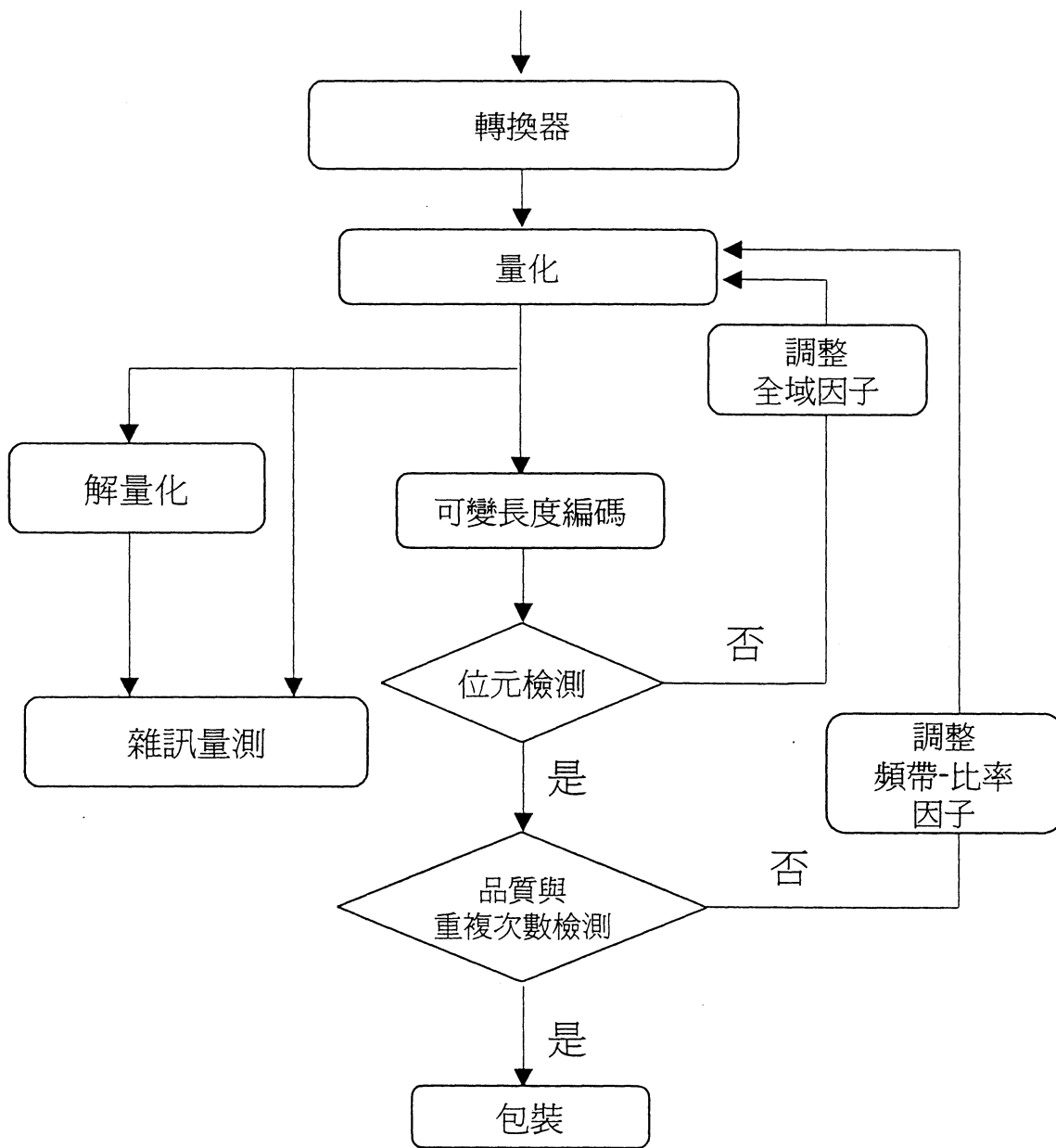


圖 2

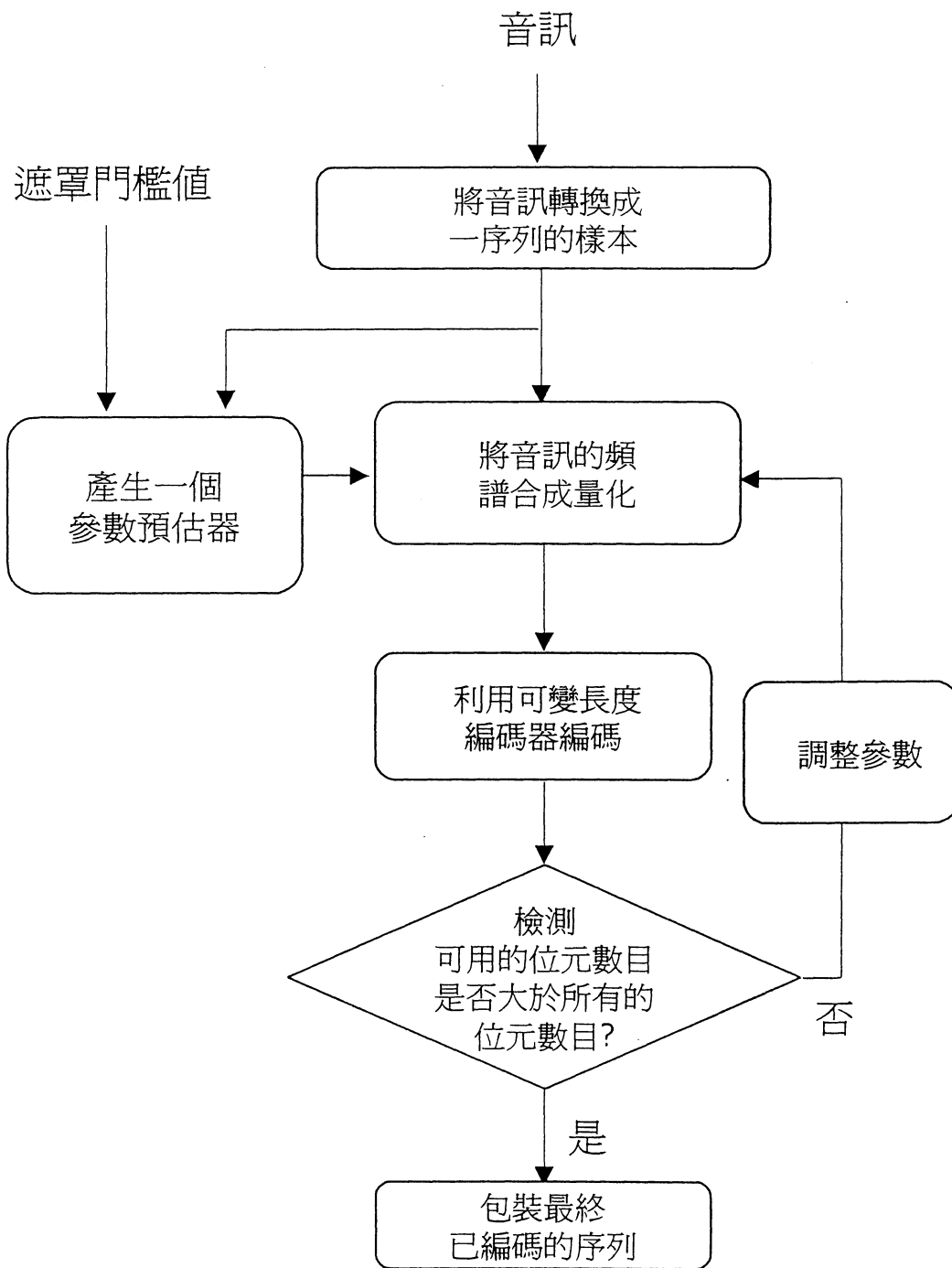


圖 3a

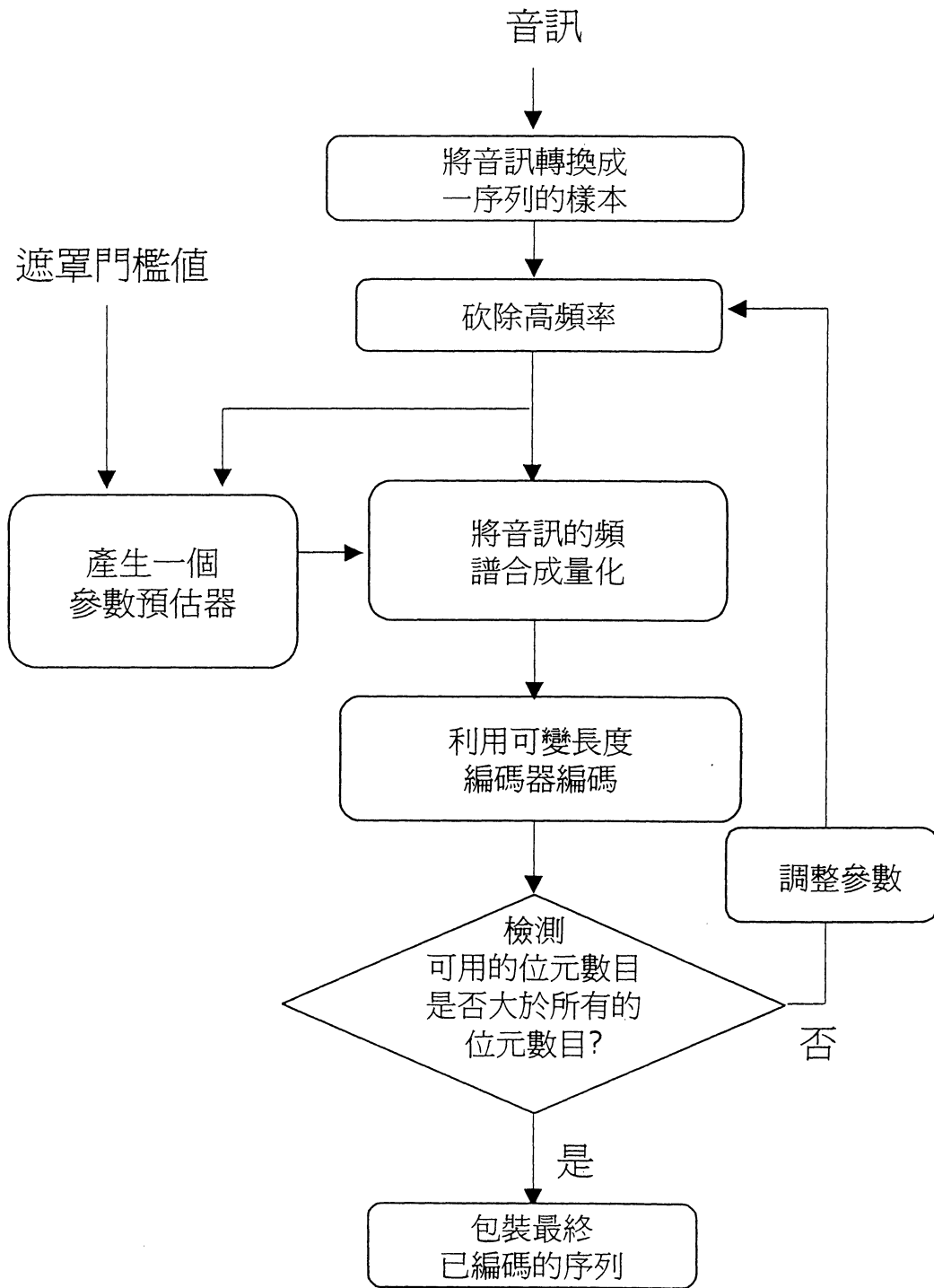
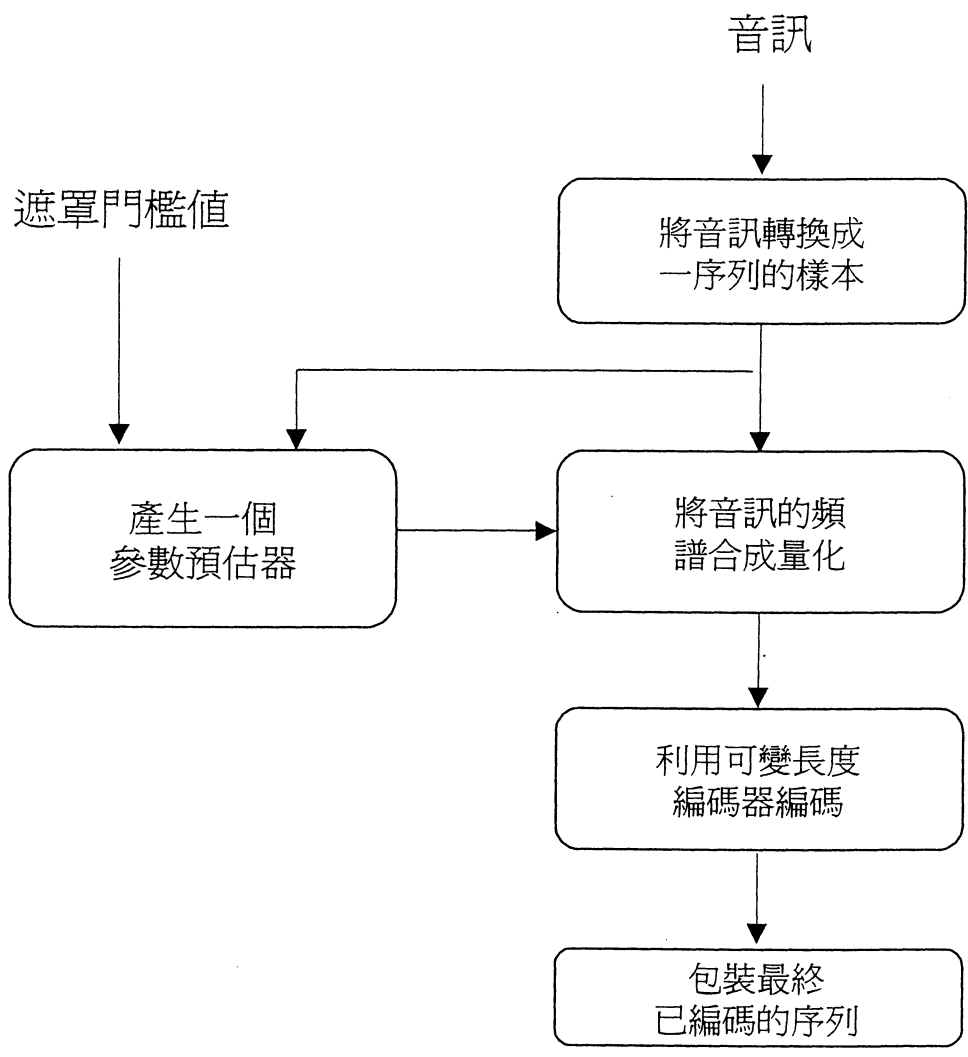


圖 3b



圖

3c

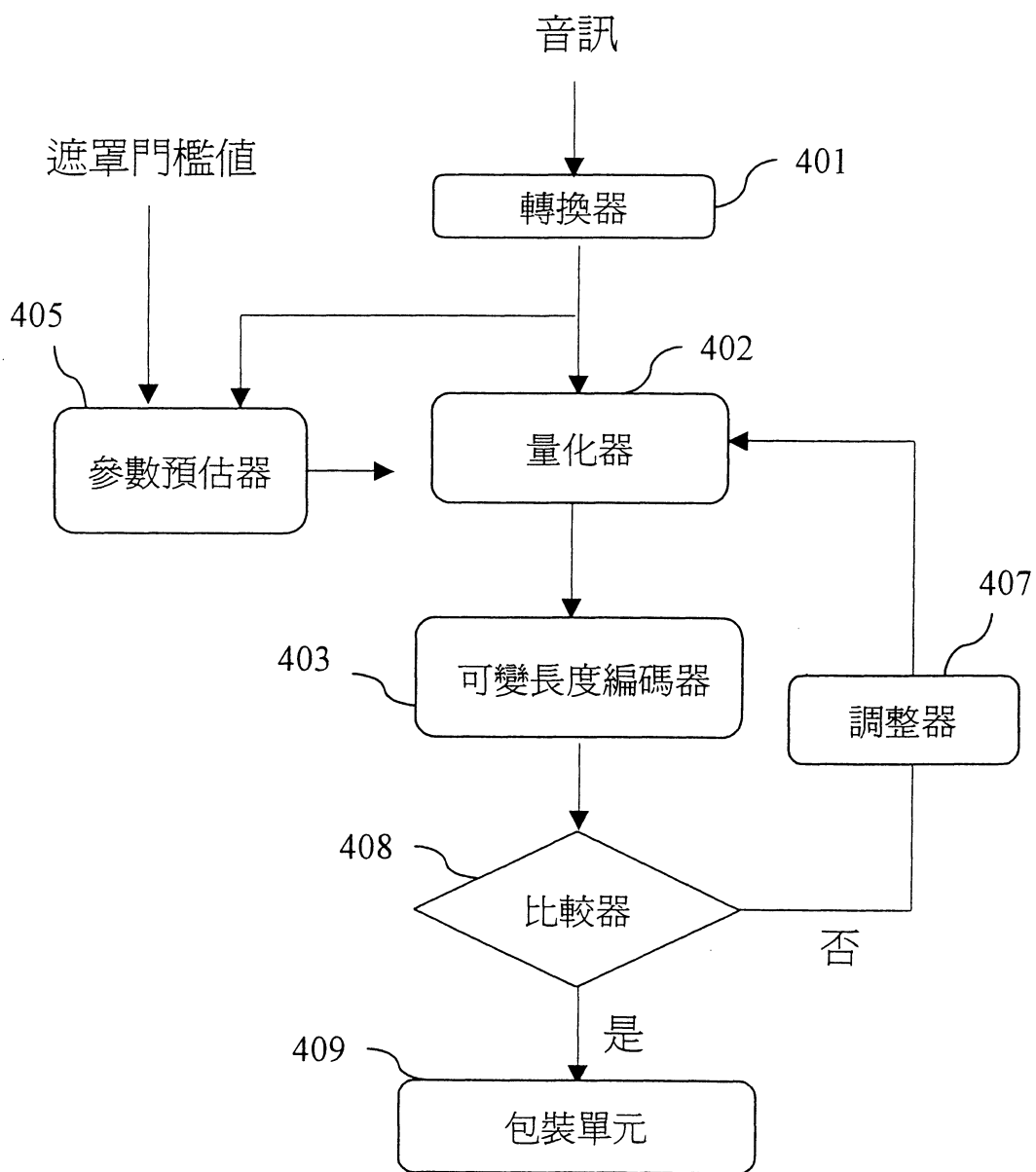


圖 4a

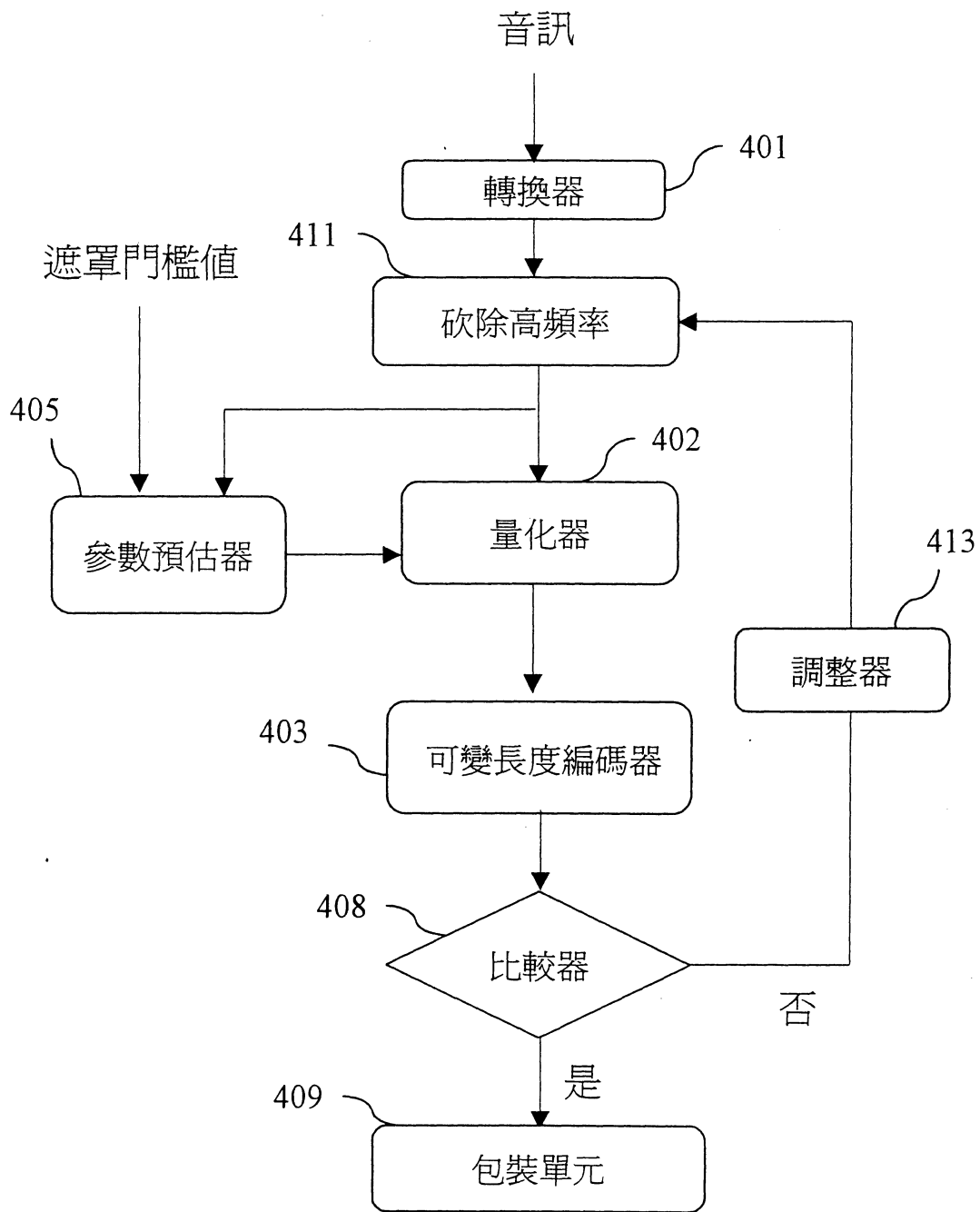


圖 4b

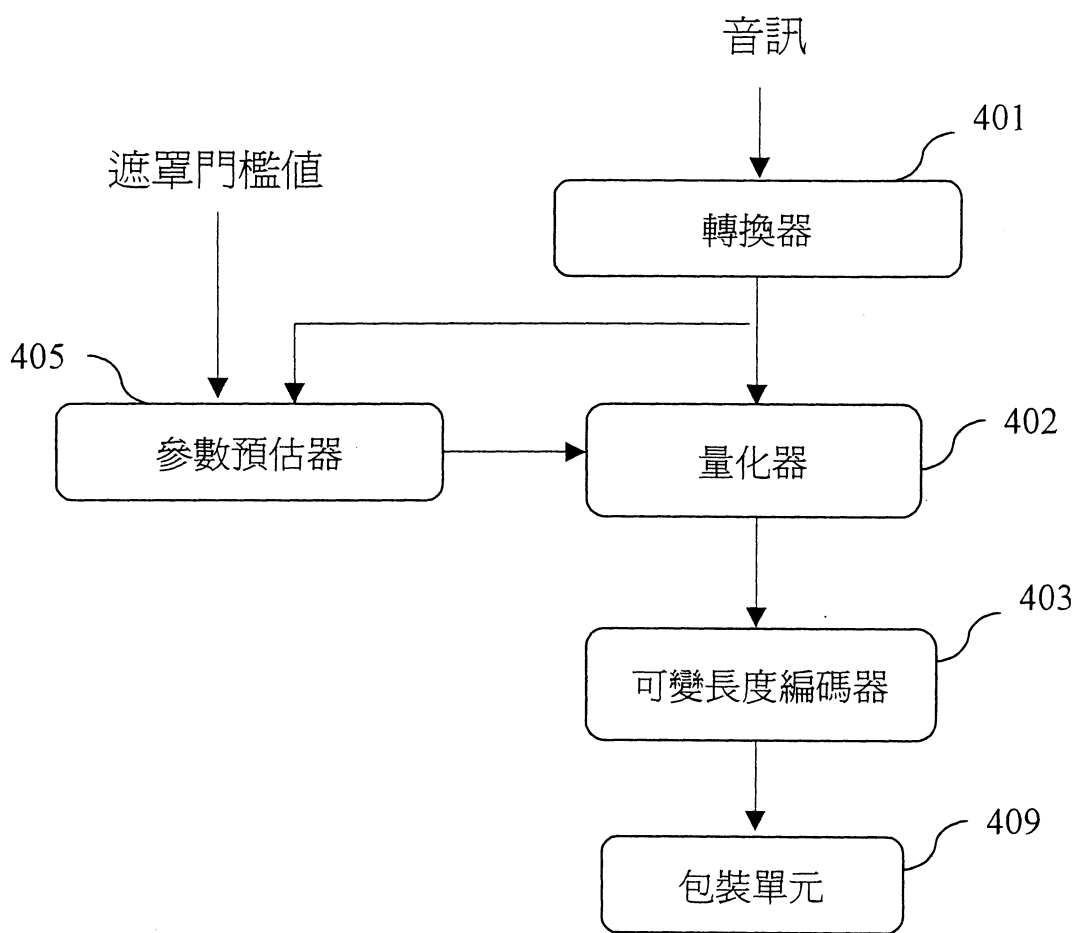


圖 4c



軌	1	2	3	4	5	6	7	8
ISO 中的 R(128kbps)	38.33	39.52	29.66	40.77	37.79	39.39	38.98	38.07
ISO 中的 Q(128kbps)	11.39	13.76	8.97	11.32	11.10	11.71	12.68	11.63
本發明中的 R	12.20	12.58	11.98	11.33	12.99	12.90	11.61	11.30
本發明中的 Q	0	0	0	0	0	0	0	0

圖 5

	1	2	3	4	5	6	7	8
ISO	-2.11	-2.32	-2.06	-1.48	-2.42	-2.22	-1.28	-1.67
ISO1	-1.84	-1.88	-1.58	-1.19	-2.44	-1.88	-0.99	-1.27
本發明	-1.71	-1.35	-1.87	-0.82	-1.93	-1.64	-0.94	-1.17

圖 6

號碼	項目	檔名	備註
1	Bass	Bass	4
2	Glockenspiel	Gspi	1 & 2 & 5
3	Harpsichord	Hach	1 & 2 & 4
4	Horn	Horn	4
5	Quarter	Quar	4
6	Soprano	Sopr	4
7	Trumpet	Trpt	2
8	Violoncello	Vioo	5

備註：

- 1) 暫時性逗留的：前回音敏感的，在暫時性區域裡雜訊的模糊化。
- 2) 音色結構：雜訊敏感的，粗糙不平滑的。
- 3) 自然語音(重要的音色部份和襲擊動作的組合)：歪曲敏感的，襲擊動作的模糊化。
- 4) 複雜的聲音：加強此測試的元件。
- 5) 高頻寬：加強次測試的元件，高頻率的遺漏，已程式化之調變得高頻雜訊。

圖 7