

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※申請案號：*ab111338*

※申請日期：*96.3.30*

※IPC 分類：*H03G 3/32 (2006.01)*

一、發明名稱：(中文/英文)

利用聽覺事件之聽覺場景分析與特定響度式檢測來控制音訊動態增益參數之技術

CONTROLLING DYNAMIC GAIN PARAMETERS OF AUDIO USING AUDITORY SCENE ANALYSIS

AND SPECIFIC-LOUDNESS-BASED DETECTION OF AUDITORY EVENTS

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

杜比實驗室特許公司 / DOLBY LABORATORIES LICENSING CORPORATION

代表人：(中文/英文)

張莉妮 / CHEUNG, LILY S.

住居所或營業所地址：(中文/英文)

美國加州舊金山市·帕翠洛街 100 號

100 Potrero Avenue, San Francisco, CA 94103, U.S.A

國籍：(中文/英文)

美國 / U.S.A.

三、發明人：(共 2 人)

姓名：(中文/英文)

1. 克洛奇特 布雷特 G. / CROCKETT, BRETT GRAHAM

2. 西費爾特 亞倫 J. / SEEFELDT, ALAN JEFFREY

國籍：(中文/英文)

1. 美國 / U.S.A.

2. 美國 / U.S.A.

四、聲明事項：

主張專利法第二十二條第二項 第一款或 第二款規定之事實，其事實發生日期為：。

申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

有主張專利法第二十七條第一項國際優先權：

1. 美國、 2006/04/27、 60/795,808

無主張專利法第二十七條第一項國際優先權：

主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

主張專利法第三十條生物材料：

須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

九、發明說明：

【發明所屬之技術領域】

發明領域

本發明係論及音訊動態範圍控制之方法和裝置，其中，一個音訊處理裝置，可分析一個音訊信號，並且可改變該音訊之位準、增益、或動態範圍，以及上述音訊增益和動態處理之所有或某些參數在產生上，係為一個聽覺事件函數。本發明亦論及一些用以實現該等方法或用以控制該等裝置之電腦程式。

10 本發明亦論及一些用以使用聽覺事件之特定響度式檢測的方法和裝置。本發明亦論及一些用以實現該等方法或用以控制該等裝置之電腦程式。

【先前技術】

發明背景

15 音訊之動態處理

自動增益控制(AGC)和動態範圍控制(DRC)之技術係屬習見，以及係許多音訊信號路徑之常見元件。在抽象之意義中，以上兩種技術係使用某種方法，來測量一個音訊信號之位準，以及接著修飾該信號之增益，而達一個屬上述測得位準之函數的數量。在一種線性1:1動態處理系統中，該輸入音訊並不會被處理，以及其輸出音訊信號，理想上係與該輸入音訊信號相匹配。此外，若在此有一種音訊動態處理系統，可自動測量上述輸入信號之特徵，以及可使用該測量值來控制輸出信號，且若該輸入信號之位準

升高6分貝，以及其輸出信號係經過處理，使其位準僅升高3分貝，則該輸出信號係相對於該輸入信號壓縮2:1之比率。國際出版品編號WO 2006/047600 A1 (“Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal” (Alan Jeffrey Seefeldt 5 所著之“計算及調整一個音訊信號之感知響度和／或感知頻譜平衡”)，提供了五種基本類型之音訊動態處理的詳細總結：壓縮、限定、自動增益控制(AGC)、擴展、和閘控。聽覺事件和聽覺事件檢測

10 將聲音分割成一些可清楚分開被感知之單元或節段，有時係被稱為“聽覺事件分析”或“聽覺場景分析”(“ASA”)，以及該等節段有時係被稱為“聽覺事件”或“音訊事件”。“聽覺場景分析”之詳盡討論，係由Albert S. Bregman闡明在其著作Auditory Scene Analysis--The 15 Perceptual Organization of Sound(聽覺場景分析-音訊感知組織)，麻省理工學院，1991年，第四版，2001年，第二MIT出版社平裝版。此外，1999年十二月14日頒給Bhadkamkar, et al之美國專利申請序列號碼第6,002,776號，引述回溯自 1976年名為“prior art work related to sound separation by 20 auditory scene analysis.”(與藉由聽覺場景分析之聲音分離有關的先存技藝式工作”)。然而，該Bhadkamkar, et al專利，阻礙了聽覺場景分析之實際使用，其結論是“一些涉及聽覺場景分析之技術，雖然就科學之觀點而言，屬一個關注人類聽覺處理之模型的課題，目前係過度仰仗電腦計算之技

能，以及在專業上係被視為有關聲音分離之實用技術，直至基礎進步被完成為止”。

一種用以識別聽覺事件之有用方法，係由Crockett和Crocket et al 闡明在下文條列於“*Incorporation by Reference.*”(引置條款)標題下之各種專利申請案和論文中。依據此等文獻，一個音訊信號，係藉由檢測頻譜複合組成(一個頻率函數之幅度)中相對於時間之變化，使分割成一些各朝向被感知為清楚分開的聽覺事件。此在完成上舉例而言，可能係藉由計算上述音訊信號之連續時間區段的頻譜內容中之差異，以及藉由在該等連續時間區段間之頻譜內容中的差異，超過一個臨界值時，識別一個聽覺事件邊界，而作為該等連續時間區段間之邊界。替代選擇的是，其可能代替或加上頻譜複合組成中相對於時間之變化，而計算上述相對於時間之幅度中的變化。

在其仰仗電腦計算技能為最少之實現體中，程序係將音訊分割成時間節段，其係藉由分析整個頻帶(全頻寬音訊)或大體上之整個頻帶(在實際之實現體中，常被採用的是在頻譜之端部處的頻帶極限濾波運作)，以及將最大之權量給最大響度之音訊信號成分。此種解決方案係利用一種心理聲學現象，其中，在較小之時間標度(20毫秒(ms)和更小)下，人耳在一個給定之時間下，可能傾向於集中注意在一個單一聽覺事件上面。此係意味雖然可能會有多重事件在同一時刻下發生，一個成分傾向於在知覺上最為顯著，以及可能被個別處理，而當作其為唯一發生之事件。利用此

一效應，亦可容許該聽覺事件檢測，按被處理之音訊的複雜性來衡量。舉例而言，若上述被處理之輸入音訊信號，係一種獨奏樂器，該等被識別之音訊事件，將有可能是正在演奏之個別音符。同樣，就一個輸入聲音信號而言，語音之個別成分，舉例而言，母音和子音，將有可能被識別為個別之音訊元素。隨著上述音訊之複雜性的增加，諸如一個鼓聲或多重樂器和聲音之音樂，該聽覺事件檢測，可識別任一給定時刻下之“最顯著”(亦即，響度最大)的音訊元素。

10 在較大之電腦計算複雜性的代價下，其程序亦可能係考慮離散副頻帶(固定或動態判定式或固定或動態兩者之判定式副頻帶)而非全頻寬內之頻譜複合組成中相對於時間的變化。此替代選擇方案，係考慮不同副頻帶內超過一個之音訊串流，而非假定在一個特定之時間下，僅有一個單一串流可被感知。

聽覺事件檢測在實現上，可能係藉由使用一個濾波器排組或時間-頻率變換運作，諸如FFT，將一個時域音訊波形，分割成一些時間間隔或區段，以及接著將每個區段內之資料，轉換成上述之頻域。每個區段之頻譜內容的幅度可能被常態化，藉以消除或降低幅度變化之效應。每個成就之頻域表示值，提供了上述特定之區段內的音訊之頻譜內容的指示值。彼等連續區段之頻譜內容係使相比較，以及一些大於一個臨界值之變化，可能被視為指示一個聽覺事件之瞬時開始或瞬時結束。

較佳的是，該頻域資料係如下文之討論使常態化。上述頻域資料需要被常態化之程度，可產生一種幅度指示值。因此，若此程度中之變化，超過一個預定的臨界值，

5 其一亦可能被視為指示一個事件邊界。頻譜變化和幅度變化所造成之件起始和結束點，可能一起進行邏輯或處理，而使任一類型之變化所造成的事件邊界可被識別。

雖然Crockett和Crockett at al之專利申請案和論文中所說明的技術，配合本發明之屬性係特別有用，其他用以識別聽覺事件和事件邊界之技術，可能被採用在本發明之屬性中。

【發明內容】

發明概要

音訊之傳統先存技藝式動態處理，係涉及使該音訊乘以時變性控制信號，後者可調整上述音訊之增益，而產生一個所希望之結果。“增益”係一個用以定標該音訊幅度之標度因數。該控制信號可能係以連續之方式來產生，或者係產生自音訊資料區段，但其通常係自上述被處理之音訊的某種測量形式導得，以及其變化率係由一些平滑濾波器來決定，後者有時係具有固定之特徵，有時係具有隨音訊動態而變化之特徵。舉例而言，彼等響應時間可能依據上述音訊之幅度或功率中的變化而做調整。一些類似自動增益控制(AGC)和動態範圍壓縮(DRC)等先存技藝式方法，並不會以任何心理聲學方式評估，增益變化可能在其間被感知為有缺損且當彼等可被施加而不會衍生聽覺之假象的時

間間隔。所以，傳統式音訊動態程序，時常會導入一些聽覺假象，亦即，上述動態處理之效應，會在音訊中導入一些不當之可察覺的變化。

聽覺場景分析，可識別彼等在知覺上離散之聽覺事件，每個事件係發生在兩個連續的聽覺事件邊界之間。一個增益變化所造成之聽覺上的缺損，可藉由確保增益在一個聽覺事件內更接近常數，以及藉由使大部份之變化限制在一個事件邊界附近，而使大幅降低。在壓縮器或擴展器之背景環境中，對音訊位準中之增加的響應(常被稱為著手處理(attack))，可能很快速且係與聽覺事件之最小期間相當或較短，但對一個降低(釋放或恢復)的響應可能較慢，以致一些本應呈現不變及逐漸衰減之聲音，可能在音訊上會呈現失真。在此種環境之下，使該增益恢復延遲，直至達次一邊界為止，或者在一個事件期間，使增益之變化率減緩，係屬非常有利。就自動增益控制施加而言，其中，上述音訊之中期至長期位準或響度係使常態化，以及著手處理和釋放時間兩者，因而比起一個聽覺事件之最小期間可能會很長，在事件期間使變化延遲或使增益中之變化率減緩，直至增加及降低增益兩者有關之次一事件邊界為止，係屬有利。

依據本發明之一個屬性，一個音訊處理系統，可接收一個音訊信號，以及可分析並變更上述音訊之增益和／或動態範圍的特徵。上述音訊之動態範圍修飾，經常係受控於一個動態處理系統之參數(著手處理和釋放時間、壓縮

比、等等)，後者針對上述動態處理所導入之知覺假象，係具有顯著之效應。上述音訊信號中相對於時間之信號特徵中的變化會被檢測，以及被識別為聽覺事件邊界，而使連續邊界間之一個音訊節段，構成上述音訊信號中之一個聽

5 覺事件。該等被關注之聽覺事件的特徵，可能包括該等事件類似知覺強度或期間等之特徵。某些該等一個或多個動態處理參數在產生上，至少係部份響應聽覺事件和／或一些與該等聽覺事件邊界相關聯之信號特徵中的變化程度。

典型地，一個聽覺事件，係傾向於被感知為清楚分開

10 之音訊節段。信號特徵的一個有用之度量，係包括該音訊之頻譜內容的度量，舉例而言，正如上述被引用之Crockett和Crockett et al文件中所說明者。所有或某些該等一個或多個音訊動態處理參數在產生上，可能至少部份響應一個或多個聽覺事件之存在與否和特徵。一個聽覺事件邊界，可能

15 能被識別為信號特徵中相對於超過一個臨界值之時間變化。或者，所有或某些該等一個或多個參數在產生上，可能至少部份響應一些與該等聽覺事件邊界相關聯之信號特徵中的變化程度之連續性度量。雖然本發明之屬性，原則上可能在類比域和／或數位域中被實現，實際之實現體，

20 有可能是在數位域中被實現，其中，每個音訊信號，係由個別之樣本或資料區段內之樣本來表示。在此一情況中，該等信號特徵可能是一個區段內之音訊的頻譜內容，彼等信號特徵中相對於時間之變化的檢測，可能是逐區段之音訊的頻譜內容中之變化的檢測，以及彼等聽覺事件時間起

始和停止邊界，各係與資料之區段的邊界相重合。理應注意到的是，就以逐樣本為基礎來執行動態增益改變之更傳統的情形而言，上述所說明之聽覺場景分析，可能以逐區段為基礎來執行，以及所成就之聽覺事件資訊，係被用來

5 執行逐樣本所施加之動態增益改變。

藉由使用聽覺場景分析之結果來控制關鍵性音訊動態處理參數，便可能達成動態處理所導入之聽覺的假象之顯著降低。

本發明提出了兩種用以執行聽覺場景分析之方法。第一種方法可執行頻譜分析，以及可藉由識別頻譜內容中之變化，來識別一些被用來控制該等動態增益參數之可察覺的音訊事件之位置。第二種方法可將該音訊變換成一個知覺響度域(其可能比第一種方法提供更多之心理聲學相關資訊)，以及可識別一些隨繼被用來控制該等動態增益參數

10 之聽覺事件的位置。理應注意到的是，第二種方法需要的是，上述之音訊處理，應知道絕對的聽覺再現位準，彼等在某些實現體中，可能係屬不可能。提出兩種聽覺場景分析之方法，可容許使用一些為考慮絕對再現位準而可能或

15 可能不被校準之程序或裝置，來實現ASA-控制式動態增益

20 修飾。

本發明之屬性在本說明書中，係在一個包括其他發明之屬性的音訊動態處理環境中做說明。此類其他發明係說明在本專利申請案之業主，Dolby Laboratories Licensing Corporation(杜比實驗室授權公司)，的各個待審之美國和國

際專利申請案中，該等專利申請案係指明在本說明書中。

圖式簡單說明

第1圖係一個可顯示用以執行聽覺場景分析之處理步驟的範例之流程圖；

- 5 第2圖係顯示一個在執行上述聽覺場景分析時針對音訊進行區段處理、窗取樣、及執行該DFT之範例；

第3圖係在一個流程圖或功能方塊圖之性質中顯示平行處理，其中，音訊係被用來識別聽覺事件並識別該等聽覺事件之特徵，而使該等事件和彼等之特徵，被用來修飾
10 該等動態處理參數；

第4圖係在一個流程圖或功能方塊圖之性質中顯示處理，其中，音訊係僅被用來識別聽覺事件，以及該等事件之特徵，係導自該音訊事件檢測，而使該等事件和彼等之特徵，被用來修飾該等動態處理參數；

- 15 第5圖係在一個流程圖或功能方塊圖之性質中顯示處理，其中，音訊係僅被用來識別聽覺事件，以及該等事件之特徵，係導自該音訊事件檢測，而僅使該等聽覺事件之特徵，被用來修飾該等動態處理參數；

第6圖係顯示一組可近似化基於ERB標度之臨界頻帶
20 的理想化之聽覺濾波器特徵響應。其水平標度為赫茲(Hz)下之頻率，以及其垂直標度為分貝(dB)下之位準；

第7圖係顯示ISO 226之等響度曲線圖，其水平標度為赫茲(Hz)(基底10之對數標度)下之頻率，以及其垂直標度為分貝(dB)下之聲壓位準；

第8a-c圖係顯示一個音訊動態範圍壓縮器之理想化的輸入／輸出特徵和輸入增益特性；

第9a-f圖係顯示一個使用聽覺事件來控制一個傳統式動態範圍控制器(DRC)之數位實現體中的釋放時間之聽覺事件的範例，其中，該增益控制係導自該信號之均方根(RMS)功率；

第10a-f圖係顯示一個使用聽覺事件來控制一個傳統式動態範圍控制器(DRC)之數位實現體中的釋放時間之聽覺事件的範例，其中，該增益控制係導自一個不同於第9圖中所使用之替代信號的信號之均方根(RMS)功率；而

第11圖則係描述在一個響度域動態處理系統中施加AGC緊跟DRC有關之理想化的AGC和DRC曲線之適當集合。此種組合之目標，係使所有處理之音訊，具有大約相同之感知響度，同時仍能維持至少某些原始音訊動態。

15 **【實施方式】**

較佳實施例之詳細說明

聽覺場景分析(原形非響度域方法)

依據本發明之一個屬性的實施例中，聽覺場景分析，可能係由第1圖之一部分所顯示的四個一般性處理步驟所組成。第一步驟1-1(“執行頻譜分析”)，可取得時域聲頻信號，可將其分割成一些區段，以及可計算每個區段有關之頻譜廓線或頻譜內容。頻譜分析可將該音訊信號變換成短期頻域。此在執行上亦可能使用任何濾波器組，彼等或基於變壓器，或基於帶通濾波器組，以及係在線性或翹曲性

頻率空間(諸如巴克量度或臨界頻帶，彼等更接近人耳之特性)。以任何一個濾波器組，在時間與頻率之間，存在著一種折衷選擇。較大之時間解析度，和因而較短之時間間隔，將導致較低之頻率解析度。較大之頻率解析度，和因而較窄之子頻帶，將導致較長之時間間隔。

第1圖中以概念例示之第一步驟，可計算上述音訊信號之連續時間節段的頻譜內容。在一個實際之實施例中，上述ASA區段之尺度，可能來自上述輸入音訊信號之任何樣本數目，不過，512個樣本可提供時間和頻率解析度之良好折衷選擇。在第二步驟1-2中，逐區段之頻譜內容中的差異會被決定(“執行頻譜廓線差異測量”)。因此，該第二步驟可計算上述音訊信號的連續時間節段之間的頻譜內容中之差異。誠如上文所討論，一個被感知之聽覺事件的開始和結束之強效指示符，係被認為屬頻譜內容中之一個改變。在第三步驟1-3(“識別聽覺事件邊界之位置”)中，當一個頻譜廓線區段與次一區段之間的頻譜差異，大於一個臨界值時，該區段邊界係被採用為一個聽覺事件邊界。彼等連續邊界之間的音訊節段，構成了一個聽覺事件。因此，該第三步驟，可在該等連續邊界之間的頻譜廓線內容中之差異，超過一個臨界值時，設定連續時間節段之間的聽覺事件邊界，因而界定出聽覺事件。在此一實施例中，彼等聽覺事件邊界，可界定出一些聽覺事件，使彼等具有之長度，為一個頻譜廓線區段之整數倍數，其最小之長度為一個頻譜廓線區段(此範例中為512個樣本)。大體而言，事件邊界

並不必受到如此之限制。就本說明書所討論之實際實施例的替代方案而言，上述輸入區段之尺度可能有變化，舉例而言，而使基本上為一個聽覺事件之尺度。

緊接該等事件邊界之識別，如步驟1-4中所示，該聽覺事件之關鍵特徵會被識別出。上述音訊之交疊或非交疊節段，可能被窗取樣，以及被用來計算上述輸入音訊之頻譜廓線。交疊可對聽覺事件之位置，造成較細緻之解析度，以及亦可使其較不可能錯過一個事件，諸如一個短期暫態。然而，交疊亦會增加電腦計算之複雜性。因此，交疊可能會被省略。第2圖係顯示被窗取樣並以離散傅立葉變換使(DFT)轉換成頻域之非交疊的N個樣本區段之概念表示。每個區段可能被窗取樣，以及使轉換成頻域，諸如藉由使用該DFT，較佳的是，考為速率而使實現為一個快速傅立葉變換(FFT)。

以下諸變數可能被用來計算上述輸入區段之頻譜廓線：

M = 一個被用來計算頻譜廓線之區段中的窗取樣之樣本的數目

P = 頻譜計算交疊之樣本數目

通常，以上變數可能使用任何之整數。然而，若 M 被設定使等於一個2之幂方，而使標準FFT可能被用於該頻譜廓線計算，則其實現體將更有效率。在上述聽覺場景分析程序之實際實施例中，該等列舉之參數可能被設定為：

$M = 512$ 個樣本(或44.1仟赫下之11.6 ms)

P = 0個樣本(非交疊)

以上列舉之值係以實驗來決定，以及通常被發現可以足夠之準確度，來識別聽覺事件之位置和期間。然而，設定P之值為256個樣本(50%交疊)而非零個樣本(非交疊)，已被發現用以識別某些難以發現之事件係屬有用。雖然許多不同類型之取樣窗，可能被用來極小化取樣窗所致之頻譜假象，該等頻譜廓線計算中所使用之窗取樣，M-點 Hanning(漢寧窗)、Kaiser-Bessel(凱塞-貝塞爾窗)、或其他適當且較佳之非方形窗。上文所指明之值和漢寧窗類型，係在大量之實驗分析已顯示可橫跨寬廣範圍之音訊素材提供優異的結果之後被選定。處理一些具有絕大多數之低頻內容的音訊信號，較佳的是非方形取樣窗。方形取樣窗會產生一些可能造成事件之不正確的檢測之頻譜假象。不同於某些編碼器／解碼器(codec)施加，其中，總體交疊／相加程序勢必要提供一個固定之位準，此種限制條件並不適用，以及該取樣窗可能就一些類似其之時間／頻率解析度和止帶抑制(stop-band rejection)等特性，來加以選擇。

在步驟1-1(第1圖)中，每個M-取樣區段之頻譜在計算上，可能藉由以一個M-點 Hanning(漢寧窗)、Kaiser-Bessel(凱塞-貝塞爾窗)、或其他適當之取樣窗，來窗取樣該資料，使用一個M-點快速傅立葉變換，使轉換成頻域，以及計算該等複數FFT係數之幅度。此成就之資料係使常態化，而使其最大之幅度被設定為“一”，以及該等M個數目之常態化陣列，係被轉換至對數域。該資料亦可能以其

他度量使常態化，諸如該資料之平均幅度值或平均功率值。該陣列並不需要被轉換至對數域，但該項轉換可簡化步驟1-2之差異計量。此外，該對數域更密切符合人類聽覺系統之性質。上述成就之對數域值，係具有負無限大至零之範圍。在一個實際之實施例中，該等值之範圍，可能賦有一個下限；該極限可能係使固定，舉例而言-60 dB，或是係屬頻率相依性，藉以反映低頻和極高頻下之靜音的較低可聞度。(理應注意到的是，該陣列之尺度有可能被約減至 $M/2$ ，而使FFT表示負頻率和正頻率)。

10 步驟1-2可計算相鄰區段之頻譜間的差異之計量。就每個區段而言，每個來自步驟1-1之 M (對數)頻譜係數，係使自前一區段有關之對應係數減除，以及計算此差異之幅度(忽略正負號)。此等 M 個差異接著係使加總成一個數字。此差異計量亦可能藉由使此差異除以該加總中所用之頻譜係數的數目(在此情況中為 M 個係數)，而表示為每頻譜係數之平均差異。

20 步驟1-3可藉由施加一個臨界值至上述來自步驟1-2而具有一個臨界值之差異計量的陣列，來識別彼等聽覺事件邊界之位置。當一個差異計量超過一個臨界值時。頻譜中之變化，被認為係足以傳訊一個新事件，以及該變化之區段數目，係被記錄為一個事件邊界。就上文所列之 M 和 P 的值，以及就以單位dB表示之對數值(步驟1-1中)而言，若被比較的是全幅度FFT(包括鏡射之部分)，該臨界值可被設定使等於2500，以及若被比較的是一半FFT(如上文所表明)，

該臨界值便被設定為1250。就FFT之幅度而言，該FFT代表著負頻率和正頻率，一個是另一個的鏡像)此值係由實驗來選擇，以及其提供了良好之聽覺事件邊界檢測。此參數值可能被改變來降低(增加該臨界值)或增加(降低該臨界值)

5 彼等事件之檢測值。

第1圖之程序可能更一般性地以第3、4、和5圖之等效安排來表示。在第3圖中，一個音訊信號係平行施加至一個“識別聽覺事件”功能或步驟3-1，其可將該音訊信號，分割成一些各傾向於被感知為清楚分開的聽覺事件，以及施加至一個選用之“識別聽覺事件之特徵”功能或步驟3-2。第1
10 圖之程序，可能被採用來將該音訊信號分割成一些被識別之聽覺事件和彼等之特徵，或者可能採用某些其他之適當程序。上述可能由功能或步驟3-1所決定之聽覺事件邊界的身份之聽覺事件資訊，接著係依需要被用來以“修飾動態參數”功能或步驟3-3，修飾該等音訊動態處理參數(諸如著手處理、釋放、比率、等等)。該選用性“識別特徵”功能或步
15 驟3-3，亦可接收該聽覺事件資訊。該“識別特徵”功能或步驟3-3，可能藉由一個或多個特徵，來特性化某些或所有之聽覺事件。此種特徵正如配合第1圖之程序所說明的，可能
20 包括上述聽覺事件之支配性子頻帶的身份。該等特徵亦可能包括一個或多個音訊特徵，其中舉例而言包括：該聽覺事件之功率的計量、該聽覺事件之幅度的計量、該聽覺事件之頻譜平坦度的計量、和該聽覺事件是否大體上無聲、或其他有助於修飾動態參數之特徵，以便降低或移除上述

處理之無用的聽覺假象。該等特徵亦可能包括其他類似該聽覺事件是否包含有暫態等之特徵。

第3圖之安排的替代方案，係顯示在第4和5圖中。在第4圖中，該音訊輸入信號，係無法直接施加至“識別特徵”功能或步驟4-3，但其確可接收來自“識別聽覺事件”功能或步驟4-1之資訊。第1圖之安排係此種安排之一個特定範例。在第5圖中，該等功能或步驟5-1、5-2、和5-3，係按順序排列。

此實際實施例之細節並不重要。有其他之方法，可能被採用來計算上述音訊信號之連續時間節段的頻譜內容，計算該等連續時間節段之間的差異，以及在此等連續時間分段之間的頻譜廓線內容中之差異，超過一個臨界值時，設定彼等連續時間節段之間的對應邊界處之聽覺事件邊界。

15 聽覺場景分析(新的響度域方法)

2005年十月25日提出申請之專利合作條約編號PCT/US2005/038579下的國際申請案，國際公開號WO 2006/047600 A1，Alan Jeffrey Seefeldt取名為“Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal”(計算及調整一個音訊信號之感知響度和／或感知頻譜平衡)，且姑不論其他，揭示了一種基於一個心理聲學模型之感知響度的客觀計量。該申請案係藉由參照其全文使合併進本說明書內。誠如該申請案中所說明，自一個音訊信號， $x[n]$ ，有一個激勵信號

$E[b,t]$ 計算出，其可近似化時間區段 t 期間在臨界頻帶 b 下沿內耳之基底膜的能量分佈。該激勵信號可自上述音訊信號之短時離散傅立葉變換 (STDFT) 計算如下：

$$E[b,t] = \lambda_b E[b,t-1] + (1 - \lambda_b) \sum_k |T[k]|^2 |C_b[k]|^2 |X[k,t]|^2 \quad (1)$$

5 其中， $X[k,t]$ 係表示時間區段 t 和集器 k 下之 $x[n]$ 的 STDFT。理應注意到的是，在方程式 1 中， t 係表示變換區段之離散單位中的時間，而與類似秒等之連續計量形成對比。 $T[k]$ 係表示一個模擬透過外耳和中耳之音訊的傳輸之濾波器的頻率響應，以及 $C_b[k]$ 係表示一個對應於臨界頻帶 b 之
10 位置的基底膜之頻率響應。第 6 圖係描述一個適當組之臨界頻帶濾波器響應，其中，有 40 個頻帶，如 Moore 和 Glasberg 所界定，係沿等效矩形頻寬 (ERB) 標度均勻相間。每個濾波器形狀，係藉由圓通化指數函數來說明，以及該等頻帶係使用 1 ERB 之間隔來分配。最後，方程式 1 中之平滑時間常
15 數 λ_b ，可能有利的是在選擇上使正比於頻帶 b 內之人耳響度感知積分時間。

使用等響度廓線，諸如第 7 圖中所描繪者，每個頻帶處之激勵信號，係被轉換成一個將會產生 1 kHz 下之同一感知響度的激勵位準。接著自上述轉換過之激勵信號， $E_{1kHz}[b,t]$ ，
20 透過一個壓縮式非線性化，計算出一種橫跨頻率和時間而分佈之感知響度的計量之比響度。一個此種用以計算比響度 $N[b,t]$ 之適當的函數係定為：

$$N[b,t] = \beta \left(\left(\frac{E_{1kHz}[b,t]}{TQ_{1kHz}} \right)^\alpha - 1 \right) \quad (2)$$

其中， $TQ_{1\text{kHz}}$ 係 1 kHz 下之安靜中的臨界值，以及常數 β 和 α 係選擇使與收聽實驗所收集到之響度資料的成長相匹配。理論上來說，此種自激勵信號至比響度之變換，可能以 $\Psi\{\}$ 來表示，而使：

$$5 \quad N[b,t] = \Psi\{E[b,t]\}$$

最後，以單位宋 (sone) 表示之總響度， $L[t]$ ，係藉由橫跨頻帶使比響度加總來加以計算：

$$L[t] = \sum_b N[b,t] \quad (3)$$

該比響度 $N[b,t]$ 係一個意在模擬人耳感知音訊為頻率和時間的一個函數之方式的頻譜表示式。其可捕捉靈敏度對不同頻率之變動、靈敏度對位準之變動、和頻率解析度中之變動。就此而論，其係一個與聽覺事件十分匹配之頻譜表示式。雖然在電腦計算上更複雜，橫跨頻帶比較 $N[b,t]$ 在連續時間區段之間的差異，在許多情況中，與上文所說明直接使用連續之 FFT 頻譜相比較，在感知上將可能造成聽覺事件之更準確的檢測。

在該專利申請案中，有幾種用以基於此心理聲學響度模型來修飾音訊之施加例揭示出。其中有幾種動態處理運算法，諸如 AGC 和 DRC。此等揭示之運算法，可能得利於使用聽覺事件來控制各種相關聯之參數。由於比響度已被計算出，其立即可供該等事件之檢測用。一個較佳實施例之細節，係在下文中加以討論。

有關以聽覺事件來控制音訊動態處理參數

茲提出本發明之實施例的兩個範例。其第一個說明了

使用聽覺事件來控制一個動態範圍控制器(DRC)的數位實現體中之釋放時間，其中，增益控制係導自信號之均方根(RMS)功率。第二個實施例，說明了使用聽覺事件來控制上文所說明之心理聲學響度模型的背景環境中之AGC和DRC

5 實現體的更複雜之組合的某一定之屬性。此兩種實施例係僅意在充作本發明之一個範例，以及理應瞭解的是，使用聽覺事件來控制一個動態處理運算法的參數，並非受限於下文所說明之特定細節。

動態範圍控制

10 一個DRC之被說明的數位實現體，可將一個音訊信號 $x[n]$ ，分段成一些窗分半交疊之區段，以及可就每個區段，基於該信號之局部功率的計量和一個被選定之壓縮曲線，計算出一個修飾增益。此增益係橫跨諸區段使平滑化，以及接著以每個區段相乘。此等修飾過之區段，最後係使交

15 疊相加，藉以產生修飾之音訊信號 $y[n]$ 。

理應注意到的是，雖然本說明書所說明之聽覺場景分析和DRC之數位實現體，係將時域音訊信號，分割成一些區段，藉以執行分析和處理，該DRC處理在執行上，並不需要使用區段分割。舉例而言，該聽覺場景分析在執行上，

20 可能使用如上文所說明之區段分割和頻譜分析，以及所成就之聽覺事件位置和特徵，可被用來提供控制資訊，給一個在運作上以逐一取樣之基礎的傳統式DRC實現體之數位實現體。然而，在此，聽覺場景分析所使用之同一節段結構，係供DRC採用來簡化彼等組合之說明。

繼續進行一個區段式DRC實現體之說明，上述音訊信號之交疊區段可使表示為：

$$x[n,t] = w[n]x[n+tM/2] \quad \text{就 } 0 < n < M-1 \text{ 而言} \quad (4)$$

其中，M為區段長度，以及其跳距為M/2， $w[n]$ 為其開窗， n 為區段內之取樣指數，以及 t 為區段指數(理應注意到的是，在此， t 在使用上係與方程式1中之STDFT的方式相同；其舉例而言，係以區段之離散單位而非秒來表示時間)。理想上，該開窗 $w[n]$ 係逐漸變小至兩端處的零，以及當與其本身成半交疊時係加總成一；舉例而言，其常用之
5
10 正弦開窗，係符合此等準則。

就每個區段而言，該RMS功率接著可被計算，藉以產生一種以dB/每區段為單位之功率計量 $P[t]$ ：

$$P[t] = 10 * \log_{10} \left(\frac{1}{M} \sum_{n=1}^M x^2[n,t] \right) \quad (5)$$

誠如先前所提及，此功率計量，可在被一個壓縮曲線處理之前，以快速著手處理和緩慢釋放使平滑化，但就一個替代方案而言，處理的是瞬間功率 $P[t]$ ，以及所成就之增益係使平滑化。此替代解決方案具有之優點是，可能使用一個有銳形拐點之簡單的壓縮曲線，但所成就之增益在功率行經該拐點時仍呈平滑形。將一個如第8c圖中所顯示之
15
20 壓縮曲線，表示為一個可產生一個增益之信號位準函數 F ，該區段增益 $G[t]$ 係定為：

$$G[t] = F\{P[t]\} \quad (6)$$

假定隨著信號位準之增加，該壓縮曲線係施加較大之

衰減，該增益於信號處於“著手處理模態”中時將會降低，以及在“釋放模態”中時將會逐漸。所以，一個平滑化之增益，在計算上可能係依據：

$$\bar{G}[t] = \alpha[t] \cdot \bar{G}[t-1] + (1 - \alpha[t])G[t] \quad (7a)$$

5 其中，

$$\alpha[t] = \begin{cases} \alpha_{attach} & G[t] < \bar{G}[t-1] \\ \alpha_{release} & G[t] \geq \bar{G}[t-1] \end{cases} \quad (7b)$$

以及

$$\alpha_{release} \gg \alpha_{attach} \quad (7c)$$

最後，上述以dB為單位之平滑化增益 $\bar{G}[t]$ ，係施加至
10 上述信號之每個區段，以及此被修飾之區段係使交疊相加，藉以產生經修飾之音訊：

$$y[n + tM/2] = \left(10^{\bar{G}[t]/20}\right)x[n, t] + \left(10^{\bar{G}[t-1]/20}\right)x[n + M/2, t-1]$$

就 $0 < n < M/2$ 而言 (8)

理應注意到的是，由於該等區段如方程式4中所示，已
15 與一個錐形窗分(windowed)相乘，上文所示之交疊相加合成技術，可橫跨上述處理之信號 $y[n]$ 的取樣，有效地使該等增益平滑化。因此，該增益控制信號，可接收除方程式7a中所顯示者外之平滑化。在一個逐樣本而非逐區段之方式運作的DRC之較傳統的實現體中，可能有必要的，是比方
20 程式7a中所顯示之簡單的單極濾波器更複雜之平滑化的增益，藉以避免上述處理之信號的可見聞之失真。而且，使用區段式處理，可將 $M/2$ 個樣本之固有延遲導入該系統中，以及只要與 α_{attach} 相關聯之衰減時間，接近上述之延遲，該

信號 $x[n]$ 在為避免過沖而施加該等增益之前，並不需要進一步被延遲。

第9a至9c圖係描述對一個音訊信號施加上述說明之DRC處理的結果。就此特定之實現體而言，在一個44.1 kHz之取樣率下，係使用一個 $M=512$ 之區段長度。所使用的是
5 一個與第8b圖中所示者相類似之壓縮曲線：

在相對於全標度數位超過-20 dB以上，信號係以5:1之比率衰減，以及在低於-30 dB以下，信號係以5:1之比率提昇。該增益係以一個對應於10 ms之半衰減時間的著手處理
10 係數 α_{attack} 和一個對應於500 ms之半衰減時間的釋放係數 $\alpha_{release}$ 使平滑化。第9a圖中所描繪之原始音訊信號，係由六根連續鋼琴和弦所組成，其位於樣本 1.75×10^5 附近之最後和弦，係衰減至無聲。檢查第9b圖中之增益 $\bar{G}[t]$ 的曲線圖，理應注意到的是，當該等六根和弦被演奏時，該增益係保持
15 接近0 dB。此係由於該信號能量就大部份而言，係保持在-30 dB與-20 dB之間，該DRC曲線在此區域內並不需要修飾。然而，在打擊最後一根和弦之後，該信號能量係跌至低於-30 dB以下，以及隨著該和弦之衰減，該增益將會開始升高，最終超過15 dB。第9c圖描述上述成就之修飾的音
20 訊信號，以及上述最後和弦之尾部，可見係明顯被提昇。在音訊上，上述和弦之自然低位準衰減聲音的提昇，將會建立一個極不自然之結果。本發明之目標，旨在避免與一個傳統式動態處理器相關聯之此類問題。

第10a至10c圖描述施加確實相同之DRC系統至不同的

音訊信號之結果。在此一情況中，上述信號之第一半部，係包含有一個高位準下之快速樂曲，以及接著在大約 10×10^4 之取樣下，該信號係使交換至一個第二快速樂曲，但係處於一個明顯較低之強度。檢查第6b圖中之增益，該信號在

5 第一半部期間，可見係衰減大約10 dB，以及接著在該較柔和之樂曲演奏時的第二半部期間，該增益將會升高返回至0 dB。在此一情況中，該增益係依需要而變化。該第二樂曲可能想要相對於第一樂曲而使提昇，以及該增益在變遷至第二樂曲之後應使迅速增加，使在聽覺上不引人注意。人

10 們見到了一種與有關第一信號所討論者相類似之增益行為，但該行為在此係為所想要的。所以，人們想要的是固定第一種情況而不會影響到第二種情況。使用聽覺事件來控制此種DRC系統之釋放時間，提供了此種解決方案。

在第9圖中所檢查之第一信號中，上述最後和弦之衰減

15 的提昇似乎是不自然，因為該和弦和其衰減，係被感知為一個單一聽覺事件，其完整性係期望要被維持。然而，在第二種情況中，當該增益增加之際，有許多聽覺事件發生，此意謂就任何一個個別之事件，係給予很小的變化。所以，其總體增益變化並不會引發不快。人們因而可能辯解一個

20 增益改變，應僅容許在一個聽覺事件邊界之近瞬時附近。該增益在著手處理或釋放模態中時，係可施加此一原理，但就一個DRC之大多數實際的實現體而言，該增益與事件感知之人耳瞬時清晰度相較，在著手處理模態中係移動如此迅速，以致不再需要控制。所以，該DRC增益唯有在釋

放模態中時，方可能使用事件來控制其之平滑化。

茲將說明上述釋放控制之適當行為。就定性術語而言，若有一個事件被檢測到，該增益係以上文在方程式7a中所明訂之釋放時間常數使平滑化。當時間推進通過上述
5 檢測到之事件時，以及若無後繼之事件被檢測到時，該釋放時間常數將會繼續增加，而使上述被平滑化之增益，最終被“凍結”定位。若有另一個事件被檢測到，則上述之平滑化時間常數，便會被重置至其原始之值，以及該程序將會一再重複。為調整該釋放時間，依據該等檢測到之事件
10 邊界，便可能首先產生出一個控制信號。

誠如前文所討論，事件邊界可能藉由尋找上述音訊信號之連續頻譜中的變化，來加以檢測。在此一特定之實現體中，每個交疊之區段 $x[n,t]$ 的 DFT，可能被計算來產生上述音訊信號 $x[n]$ 之 STDFT：

$$15 \quad X[k,t] = \sum_{n=0}^{M-1} x[n,t] e^{-j \frac{2\pi kn}{M}} \quad (9)$$

其次，上述連續區段的常態化對數幅度之間的差異，在計算上可能依據：

$$D[t] = \sum_k |X_{NORM}[k,t] - X_{NORM}[k,t-1]| \quad (10a)$$

其中，

$$20 \quad X_{NORM}[k,t] = \log \left(\frac{|X[k,t]|}{\max_k \{|X[k,t]|\}} \right) \quad (10b)$$

在此，橫跨集器 k 之 $|X[k,t]|$ 的最大值，係供該常態化運作使用，不過，其他之常態化因數或許可能被採用；舉例

而言，橫跨集器k之 $|X[k,t]|$ 的平均值。若該差異 $D[t]$ 超過一個臨界值 D_{\min} ，則一個事件便被視為已經發生。此外，與一個最大臨界值 D_{\max} 相比較，依據 $D[t]$ 之幅度，對此一事件可能分配一個位於零與一之間的強度。此成就之聽覺事件強度

5 信號 $A[t]$ ，可能被計算為：

$$A[t] = \begin{cases} 0 & D[t] \leq D_{\min} \\ \frac{D[t] - D_{\min}}{D_{\max} - D_{\min}} & D_{\min} < D[t] < D_{\max} \\ 1 & D[t] \geq D_{\max} \end{cases} \quad (11)$$

藉由分配一個正比於上述與該事件相關聯之頻譜變化的數量之強度給該聽覺事件，可對該動態處理達成比起一個二元事件判定更大之控制。吾等發明人業已發現到，在較強之事件期間，較大之增益變化係可被接收，以及方程
10 式11中之信號，可容許此種變數控制。

該信號 $A[t]$ 係具有一個發生在一個事件邊界之位置處的脈沖之脈沖式信號。為控制該釋放時間，該信號 $A[t]$ 可能進一步被平滑化，以使其在檢測到一個事件邊界之後平滑
15 地衰減至零。上述平滑化之事件控制信號 $\bar{A}[t]$ ，可能自 $A[t]$ 被計算出，而使依據：

$$\bar{A}[t] = \begin{cases} A[t] & A[t] > \alpha_{\text{event}} \bar{A}[t-1] \\ \alpha_{\text{event}} \bar{A}[t-1] & \text{其他} \end{cases} \quad (12)$$

在此， α_{event} 控制著上述事件控制信號之衰減的時間。第9d和10d圖係描述兩個對應之音訊信號有關的事件控制
20 信號 $\bar{A}[t]$ ，而使該平滑器之半衰減時間設定至250 ms。在第

一種情況中，一個事件邊界可見係就每一六根鋼琴和弦加以檢測，以及該事件控制信號，可在每個事件之後平滑地朝向零衰減。就該第二信號而言，許多事件在檢測上係彼此在時間上十分接近，以及該事件控制信號因而絕不會完全衰減至零。

該事件控制信號 $\bar{A}[t]$ ，如今可能被使用來改變上述用來使增益平滑化之釋放時間常數。當該控制信號等於一時，來自方程式 7a 之平滑化係數 $\alpha[t]$ ，如前文所述係等於 $\alpha_{release}$ ，以及當該控制信號等於零時，該係數係等於一，而使上述被平滑化之增益免於變化。上述之平滑化係數，係使用該控制信號，使內插在該兩極值之間，其係依據：

$$\alpha[t] = \begin{cases} \alpha_{attack} & G[t] < \bar{G}[t-1] \\ \bar{A}[t]\alpha_{release} + (1 - \bar{A}[t]) & G[t] \geq \bar{G}[t-1] \end{cases} \quad (13)$$

藉由依據上述事件控制信號之一個函數連續地內插該平滑化係數，該釋放時間係被重置至一個正比於一個事件開始時之事件強度的值，以及接著在一個事件發生後，係使平滑地增加至無限大。此增加之速率係受到被用來產生上述被平滑化之事件控制信號的係數之支配。

第 9e 和 10e 圖係顯示對照於來自方程式 13 之非事件受控式係數而以來自方程式 7b 之事件受控式係數使增益平滑化之效應。在第一種情況中，該事件控制信號，在最後鋼琴和弦之後係跌至零，藉以避免增益向上移動。結果，第 9f 圖中之對應修飾的音訊，並不會蒙受上述和弦之衰減的不自然提昇。在第二種情況中，該事件控制信號，絕不會

趨至零，以及因而上述平滑化之增益信號，透過上述事件控制之施加，幾乎不會被抑制。上述被平滑化之增益的軌線，係幾乎與第10b圖中之非事件受控式增益相同。此確實是所需要之效應。

5 基於響度之AGC和DRC

就傳統式動態處理技術之替代方案而言，其中，信號修飾係一些類似峰值或RMS功率等簡單之信號測量的一個直接函數，國際專利申請案序號第PCT/US2005/038579號，揭示了使用較早被說明為一種在其中執行動態處理之架構的心理聲學式響度模型。有幾項優點被引證。首先，測量和修飾係以宋為單位來指明，其係一種比類似峰值或RMS功率等較基本之計量更精確的響度知覺之計量。第二，該音訊可能被修飾，當該總響度改變時，上述原始音訊之被感知的頻譜平衡將可被維持。在此種方式下，對總響度之改變，比起一個所舉為例利用寬帶增益來修飾音訊之動態處理器，在知覺上係變得較不明顯。最後，該心理聲學模型本質上係屬多頻帶，以及因而該系統可輕易被配置來執行多頻帶動態處理，藉以緩解一個寬帶動態處理器相關聯之習見的交叉頻譜抽運(pumping)問題。

雖然在此種響度域中執行動態處理，已擁有幾項超越一些較傳統之動態處理的優點，該項技術可能透過使用聽覺事件控制各種參數而進一步加以改善。考慮上述包含第10a圖中所描繪之鋼琴和弦和第10b和10c圖中所顯示之相關聯的DRC的音訊節段。一個相似之DRC係可在響度域中

被執行，以及在此情況中，當最後一個鋼琴和弦之衰減的響度被提昇時，此項提昇或將會較不明顯，因為上述衰減之音符的頻譜平衡，將會在施加該項提昇時被維持住。然而，一種較佳之解決方案，是全然不提昇該衰減，以及因而上述與先前就傳統式DRC所說明在響度域中使用聽覺事件來控制著手處理和釋放時間之相同原理，係可有利地被施加。

此時所說明之響度域動態處理系統，係由AGC緊接DRC來組成。此種組合之目標，係為使所有被處理之音訊，具有大致相同之感知響度，同時仍能維持至少某些原始音訊之動態。第11圖係描繪此種施加有關之AGC和DRC曲線的適當集合。理應注意到的是，兩者曲線之輸入和輸出，在表示上係以宋為單位，因為處理之執行係在響度域中。該等AGC曲線係致力於引領該輸出音訊，使更接近某些目標位準，以及如前文所提及，係以相當緩慢之時間常數來完成此一動作。該AGC可能被視為使上述音訊之長期響度達成該目標，但在短期之基礎上，該響度可能在該目標附近變動顯著。所以，動作更快之DRC，係可能被採用來使此等變動，被限制至上述特定施加認為屬可被接收之某些範圍。第11圖係顯示此種DRC曲線，其中，該AGC目標係落於上述DRC之“零頻帶”的範圍內，該曲線之此一部分並不需要修飾。以此種曲線組合，該AGC可將上述音訊之長期響度，置於上述DRC曲線之零頻帶的範圍內，以使需要施加的是極小之快速動作的DRC。若該短期響度仍在該零

頻帶外變動，該DRC則會動作而使上述音訊之響度，朝向該零頻帶移動。就一個最後一般性音符而言，上述緩慢動作之DRC，可能被施加而使上述響度模型之所有頻帶，接收相同量之響度修飾，藉以維持上述感知之頻譜平衡，以及該快速動作之DRC施加的方式，可能容許該響度修飾橫跨頻帶而變化，藉以緩解快速動作頻帶獨立性響度修飾否則或許可能造成之交叉頻譜抽運。

聽覺事件可能被利用來控制AGC和DRC兩者之著手處理和釋放。在AGC之情況中，著手處理和釋放兩者時間，比起事件知覺之瞬時清晰度係很大，以及因而事件控制可能有利地被採用在兩者情況中。藉由該DRC，上述之著手處理係相當短，以及因而事件控制，唯有上文所說明之傳統式DRC有關的釋放方有需要。

誠如前文所討論，基於事件檢測之目的，上述被採用之響度模型相關聯的比響度頻譜，係可能被使用。一個與方程式10a和10b中者相類似之差異信號 $D[t]$ ，可能係計算自方程式2中所界定之比響度 $N[b,t]$ ，其方式如下：

$$D[t] = \sum_b |N_{NORM}[b,t] - N_{NORM}[b,t-1]| \quad (14a)$$

其中，

$$N_{NORM}[b,t] = \frac{N[b,t]}{\max_b \{N[b,t]\}} \quad (14b)$$

在此，橫跨頻帶 b 之 $|N[b,t]|$ 的最大值，係供常態化使用，不過，其他之常態化因數係可能被採用；舉例而言，橫跨頻帶之 $|N[b,t]|$ 的平均值。若該項差異 $D[t]$ 超過一個臨界值 D_{min} ，

則會有一個事件被視為業已發生。該差異信號接著可能在與方程式11和12中所顯示相同之方式中被處理，藉以產生一個被用來控制該等著手處理和釋放時間之平滑事件控制信號 $\bar{A}[t]$ 。

- 5 第11圖中所描述之AGC曲線，可能被表示為一個函數，其可採用一個響度之計量作為其輸入，以及可產生一個所希望之輸出響度：

$$L_o = F_{AGC}\{L_i\} \quad (15a)$$

該DRC曲線可能同樣被表示為：

10
$$L_o = F_{DRC}\{L_i\} \quad (15b)$$

- 就該AGC而言，該輸入響度係上述音訊之長期響度的一種計量。計算此種計量，可能藉由使用相當長之時間常數(數秒之數量級)，而使方程式3中所界定之瞬間響度平滑化。已顯示出的是，在判斷一個音訊節段之長期響度中，
- 15 人們對較大聲之部分比對較低聲之部分，提供更大之權量，以及在該平滑化運作中，可能使用一個比釋放更快速之著手處理來模擬該效應。為著手處理和釋放兩者合併事件控制，決定上述AGC修飾所使用之長期響度，因而在計算上可能依照：

20
$$L_{AGC}[t] = \alpha_{AGC}[t]L_{AGC}[t-1] + (1 - \alpha_{AGC}[t])L[t] \quad (16a)$$

其中，

$$\alpha_{AGC}[t] = \begin{cases} \bar{A}[t]\alpha_{AGCattach} + (1 - \bar{A}[t]) & L[t] > L_{AGC}[t-1] \\ \bar{A}[t]\alpha_{AGCrelease} + (1 - \bar{A}[t]) & L[t] \leq L_{AGC}[t-1] \end{cases} \quad (16b)$$

此外，一個相關聯之長期比響度頻譜可能被計算出，

其在稍後可供該多頻帶DRC使用：

$$N_{AGC}[b,t] = \alpha_{AGC}[t]N_{AGC}[b,t-1] + (1 - \alpha_{AGC}[t])N[b,t] \quad (16c)$$

在實行中，該等平滑化係數可能被選擇，而使該著手處理時間，大致為該釋放時間的一半。給定該長期響度計
5 量，接著便可能計算上述與AGC相關聯之響度修飾標度，而作為輸出響度對輸入響度之比率：

$$S_{AGC}[t] = \frac{F_{AGC}\{L_{AGC}[t]\}}{L_{AGC}[t]} \quad (17)$$

該DRC修飾如今可能在施加該AGC標度之後自該響度計算出。與其在施加該DRC曲線之前使響度之計量平滑
10 化，替代方案是施加該DRC曲線至瞬間響度，以及接著隨繼使此成就之修飾平滑化。此係與先前就平滑化傳統式DRC之增益所說明的技術相類似。此外，該DRC可能施加在一種多頻帶方式中，此意謂該DRC修飾，係每個頻帶b中之比響度 $N[b,t]$ 而非總響度 $L[t]$ 的一個函數。然而，為維持上述
15 原始音訊之平均頻譜平衡，對每個頻帶可能施加DRC，而使其所成就之修飾，具有相同之平均效應，正如施加DRC至總響度時所產生者。此在達成上可能藉由以長期總響度（在施加該AGC標度之後）對長期比響度之比率來標定每個頻帶，以及使用此值作為上述DRC函數之引數。此結果接
20 著以上述比率之倒數重新標定，藉以產生上述之輸出比響度。因此。每個頻帶中之DRC標定，在計算上可能依據：

$$S_{DRC}[b,t] = \frac{N_{AGC}[b,t]}{S_{AGC}[t]L_{AGC}[t]} F_{DRC} \left\{ \frac{S_{AGC}[t]L_{AGC}[t]}{N_{AGC}[t]} N[b,t] \right\} \quad (18)$$

該等AGC和DRC修飾，可能接著使相結合，藉以形成每頻帶一個總響度標度：

$$S_{TOT}[b,t] = S_{AGC}[t]S_{DRC}[b,t] \quad (19)$$

此總標度接著可能以快速著手處理和緩慢釋放且僅對
5 釋放施加事件控制，獨立於每個頻帶使橫跨時間而平滑化。理想之平滑化在執行上，與上述傳統式DRC之增益相類似，係以上述標慶之對數為基礎，在彼等之分貝表示式中加以平滑化，不過此非屬必要。為確使上述平滑化之總
10 標度，與每個頻帶中之比響度同步移動，該等著手處理和釋放模態，可能透過比響度本身之同時平滑化，來加以決定：

$$\bar{S}_{TOT}[b,t] = \left(\alpha_{TOT}[b,t] \log(\bar{S}_{TOT}[b,t-1]) + (1 - \alpha_{TOT}[b,t]) \log(S_{TOT}[b,t]) \right) \quad (20a)$$

$$\bar{N}[b,t] = \alpha_{TOT}[b,t] \bar{N}[b,t-1] + (1 - \alpha_{TOT}[b,t]) N[b,t] \quad (20b)$$

其中，

$$\alpha_{TOT}[b,t] = \begin{cases} \alpha_{TOTattack} & N[b,t] > \bar{N}[b,t-1] \\ \bar{A}[t] \alpha_{TOTrelease} + (1 - \bar{A}[t]) & N[b,t] \leq \bar{N}[b,t-1] \end{cases} \quad (20c)$$

最後，一個目標比響度，可能基於上述施加至原始之比響度的平滑化標度，來加以計算。

$$\hat{N}[b,t] = \bar{S}_{TOT}[b,t] N[b,t] \quad (21)$$

以及接著求增益 $G[b,t]$ 之解，以及在施加至上述原始之
20 激勵值時，可使一個比響度與上述之目標相等：

$$\hat{N}[b,t] = \Psi \{ G^2[b,t] E[b,t] \} \quad (22)$$

該等增益可能施加至上述被用來計算該激勵值之濾波

器組的每個頻帶，以及接著可能藉由使該濾波器組反轉，來產生一個被修飾之時間域音訊信號，而使產生上述被修飾之音訊。

額外之參數控制

- 5 雖然上文之討論，已集中在經由上述被處理之音訊的聽覺場景分析來控制該等AGC和DRC著手處理和釋放參數，其他重要參數，亦可能得利於受到經由ASA結果之控制。舉例而言，上述來自方程式12之事件控制信號 $\bar{A}[t]$ ，可能被用來改變上述用以動態調整音訊之增益的DRC比率參數
- 10 數之值。該比率參數，與該等著手處理和釋放時間參數相類似，可能顯著地促成動態增益調整所導入之知覺假象。
- 實現體

本發明可能在硬體或軟體或兩者之組合(例如，可程式化邏輯陣列)中加以實現。除非另有指明，該等納為本發明

15 之一部分的運算法，並非本質上與任何特定之電腦或其他之裝置相關聯。特言之，各種通用型機器均可能被使用，而使寫成之程式依據本說明書之授義內容，或者更方便的是構造更專屬性之裝置(例如，積體電路)，藉以執行所需要之方法步驟。所以，本發明可能在一個或多個電腦程式中

20 被實現，彼等可在一個或多個可程式規劃式電腦系統上面執行，後者各係包含有：至少一個處理器、至少一個資料儲存系統(包括揮發性和非揮發性記憶和/或儲存元件)、至少一個輸入裝置或接口、和至少一個輸出裝置或接口。程式碼係施加至輸入資料，藉以執行本說明書所說明之功

能，以及產生輸出資訊。此輸出資訊係在已知之方式中，施加至一個或多個輸出裝置。

每個此種程式可能在任何希望之電腦語言(包括機器語言、組合語言、或高階程序語言、邏輯語言、或物件導向程式規劃語言)被實現，藉以與一個電腦系統相通訊。在
5 任一情況下，該語言可能為一種編譯或解譯語言。

每個此種電腦程式，最好是使儲存在或下載至儲存媒體或裝置(例如，固態記憶體或媒體、或磁性或光學媒體)上面，後者可被一個通用型或專屬性可程式規劃式電腦讀
10 取，以便在該儲存媒體或裝置，被該電腦系統讀取時，配置及操作該電腦，藉以執行本說明書所說明之程序。此種原創性系統，可能亦被視為實現為一個電腦可讀取式儲存媒體，其可以一個電腦程式來配置。其中，上述如此配置之儲存媒體，可使一個電腦系統，在一種特定且預定之方
15 式中運作，藉以執行本說明書所說明之功能。

已說明的為本發明之許多實施例。然而，理應瞭解的是，有各種修飾體可能被完成，而不違離本發明之精神與
20 界定範圍。舉例而言，本說明書所說明之某些步驟，可能係呈順序獨立性，以及因而可能在不同於所說明之順序中被執行。

理應瞭解的是，本發明之其他變更形式和修飾體的實現體和其之各種屬性，將可為本技藝之專業人員所明瞭，以及本發明並非受限於此等所說明之特定實施例。所以，本發明被預期係涵蓋任何和所有落在本說明書所揭示且主

張之基本原理的真實精神和界定範圍內之修飾體、變更形式、或等價體。

藉由參照合併

以下之專利權、專利申請案、和刊物，係藉由參照彼
5 等各自之全文而使納入。

音訊動態處理

音訊工程師參考書，Michael Talbot-Smith所編輯，第二版，限制器和壓縮器，Alan Tutton, 2-1492-165。Focal印刷廠，教育和專業出版公司，1999年。

10 檢測及使用聽覺事件

美國專利權申請編號 10/474,387，Brett Graham Crockett之“High Quality Time-Scaling and Pitch-Scaling of Audio Signals”(音訊信號之高品質時間標定和音調標定)，2004年六月24日出版，美國公告專利資料庫編號 US
15 2004/0122662 A1。

美國專利權申請編號 10/478,398，Brett G. Crockett et al 之“Method for Time Aligning Audio Signals Using Characterizations Based on Auditory Events”(使用基於聽覺事件之特徵使音訊信號在時間上校準的方法)，2004年七月
20 29日出版，美國公告專利資料庫編號 US 2004/0148159 A1。

美國專利權申請編號 10/478,538，Brett G. Crockett之“Segmenting Audio Signals Into Auditory Events”(使音訊信號分段成聽覺事件)，2004年八月26日出版，美國公告專利資料庫編號 2004/0165730 A1。本發明之屬性，提供了一種

除Crockett之上述申請案中所揭示者外的檢測聽覺事件之方法。

美國專利權申請編號10/478,397, Brett G. Crockett et al
之“Comparing Audio Using Characterizations Based on
5 Auditory Events”(使用基於聽覺事件之特徵來比較音訊),
2004年九月2日出版, 美國公告專利資料庫編號US
2004/0172240 A1。

國際專利合作條約公開專利資料庫編號PCT/US
05/24630, 2005年七月13日提交, Michael John Smithers之
10 標題“Method for Combining Audio Signals Using Auditory
Scene Analysis”(使用聽覺場景分析結合音訊信號之方
法), 2006年三月9日出版為WO 2006/026161。

國際專利合作條約公開專利資料庫編號PCT/US
2004/016964, 2004年五月27日提交, Alan Jeffrey Seefeldt et
15 al之標題“Method, Apparatus and Computer Program for
Calculating and Adjusting the Perceived Loudness of an
Audio Signal”(計算及調整一個音訊信號之感知響度的方
法、裝置、和電腦程式), 2004年十二月23日出版為WO
2004/111994 A2。

20 國際專利合作條約公開專利資料庫編號PCT/US
2005/038579, 2005年十月25日提交, Alan Jeffrey Seefeldt
之標題“Calculating and Adjusting the Perceived Loudness
and/or the Perceived Spectral Balance of an Audio Signal”
(計算及調整一個音訊信號之感知響度和感知頻譜平衡), 以

及出版為國際出版品編號WO 2006/047600。

Brett Crockett和Michael Smithers之“A Method for Characterizing and Identifying Audio Based on Auditory Scene Analysis,” (依據聽覺場景分析來特徵化及識別音訊之方法)，2005年五月28-31日在巴賽隆納之第118屆大會，Audio Engineering Society Convention Paper 6416(音訊工程協會大會用論文6416)。

Brett Crockett之“High Quality Multichannel Time Scaling and Pitch-Shifting using Auditory Scene Analysis,” (使用聽覺場景分析之高品質多通道時間標定和音調偏移)，2003年十月在紐約之音訊工程協會大會用論文5948。

Alan Seefeldt et al之. “A New Objective Measure of Perceived Loudness” (感知響度之新的客觀計量)，2004年十月28日在舊金山之音訊工程協會大會用論文6236。

Handbook for Sound Engineers, The New Audio Cyclopedia,(聲音工程師手冊，新音訊百科全書)，由Glen M. Ballou編輯，第二版，動態學，850-851。Focal印刷廠，Butterworth-Heinemann標記，1998年。

Audio Engineer's Reference Book(音訊工程師參考書)，由Michael Talbot-Smith編輯，第二版，第2.9節(Alan Tutton之“Limiters and Compressors”(限制器和壓縮器))，第2.149-2.165頁，Focal印刷廠，教育和專業出版公司，1999年。

【圖式簡單說明】

第1圖係一個可顯示用以執行聽覺場景分析之處理步驟的範例之流程圖；

第2圖係顯示一個在執行上述聽覺場景分析時針對音訊進行區段處理、窗取樣、及執行該DFT之範例；

5 第3圖係在一個流程圖或功能方塊圖之性質中顯示平行處理，其中，音訊係被用來識別聽覺事件並識別該等聽覺事件之特徵，而使該等事件和彼等之特徵，被用來修飾該等動態處理參數；

10 第4圖係在一個流程圖或功能方塊圖之性質中顯示處理，其中，音訊係僅被用來識別聽覺事件，以及該等事件之特徵，係導自該音訊事件檢測，而使該等事件和彼等之特徵，被用來修飾該等動態處理參數；

15 第5圖係在一個流程圖或功能方塊圖之性質中顯示處理，其中，音訊係僅被用來識別聽覺事件，以及該等事件之特徵，係導自該音訊事件檢測，而僅使該等聽覺事件之特徵，被用來修飾該等動態處理參數；

第6圖係顯示一組可近似化基於ERB標度之臨界頻帶的理想化之聽覺濾波器特徵響應。其水平標度為赫茲(Hz)下之頻率，以及其垂直標度為分貝(dB)下之位準；

20 第7圖係顯示ISO 226之等響度曲線圖，其水平標度為赫茲(Hz)(基底10之對數標度)下之頻率，以及其垂直標度為分貝(dB)下之聲壓位準；

第8a-c圖係顯示一個音訊動態範圍壓縮器之理想化的輸入／輸出特徵和輸入增益特性；

第9a-f圖係顯示一個使用聽覺事件來控制一個傳統式動態範圍控制器(DRC)之數位實現體中的釋放時間之聽覺事件的範例，其中，該增益控制係導自該信號之均方根(RMS)功率；

- 5 第10a-f圖係顯示一個使用聽覺事件來控制一個傳統式動態範圍控制器(DRC)之數位實現體中的釋放時間之聽覺事件的範例，其中，該增益控制係導自一個不同於第9圖中所使用之替代信號的信號之均方根(RMS)功率；而

- 10 第11圖則係描述在一個響度域動態處理系統中施加AGC緊跟DRC有關之理想化的AGC和DRC曲線之適當集合。此種組合之目標，係使所有處理之音訊，具有大約相同之感知響度，同時仍能維持至少某些原始音訊動態。

【主要元件符號說明】

(無)

五、中文發明摘要：

在一個揭示之屬性中，動態增益修飾，係至少部份響應聽覺事件和／或該等聽覺事件邊界相關聯之音訊信號中的改變程度，使施加至音訊信號。在另一個屬性中，一個音訊信號，係藉由比較該音訊信號之連續時間區段間的比響度(specific loudness)中之差異，使分割成一些聽覺事件。

六、英文發明摘要：

In one disclosed aspect, dynamic gain modifications are applied to an audio signal at least partly in response to auditory events and/or the degree of change in signal characteristics associated with said auditory event boundaries. In another aspect, an audio signal is divided into auditory events by comparing the difference in specific loudness between successive time blocks of the audio signal.

十、申請專利範圍：

1. 一種音訊處理方法，其中，一個處理器，可接收一個輸入通道，以及可產生一個輸出通道，後者係藉由施加動態增益修飾至該輸入通道來產生，此種方法包含之步驟有：

檢測上述音訊輸入通道中相對於時間之信號特徵中的變化；

將上述音訊輸入通道中相對於時間之信號特徵中的變化，辨識為聽覺事件邊界，其中，連續邊界間之音訊節段，構成了上述通道中之一個聽覺事件；以及

至少部份地響應聽覺事件和／或一些與該等聽覺事件邊界相關聯之信號特徵中的變化程度，來產生此種音訊動態增益修飾方法之所有或某些一個或多個參數。

2. 如申請專利範圍第1項之方法，其中，一個聽覺事件，係一個傾向於被感知為清楚分開的音訊節段。
3. 如申請專利範圍第1或2項之方法，其中，該信號特徵係包括上述音訊之頻譜內容。
4. 如申請專利範圍第1或2項之方法，其中，該信號特徵係包括上述音訊之知覺響度。
5. 如申請專利範圍第1-4項任一項之方法，其中，所有或某些一個或多個參數在產生上，係至少部份地響應一個或多個聽覺事件之存在與否。
6. 如申請專利範圍第1-4項任一項之方法，其中，該辨識動作可將信號特徵相對於時間中超過一個臨界值之變

化，識別為一個聽覺事件邊界。

7. 如申請專利範圍第1-4項任一項之方法，其中，該聽覺事件邊界，可能藉由一個函數來加以修飾，藉以建立一個可用來修飾該等音訊動態增益修飾參數之控制信號。
8. 如申請專利範圍第1-4項任一項之方法，其中，所有或某些一個或多個參數在產生上，係至少部份地響應一些與該等聽覺事件邊界相關聯之信號特徵中的變化程度之連續計量。
9. 一種可被配接來執行申請專利範圍第1至8項任一項之方法的裝置。
10. 一種電腦程式，其係儲存在一個電腦可讀取式媒體上面，而可使一個電腦控制如申請專利範圍第9項之裝置。
11. 一種電腦程式，其係儲存在一個電腦可讀取式媒體上面，而可使一個電腦控制如申請專利範圍第1至8項任一項之方法。
12. 一種用以將音訊信號分割成一些聽覺事件之方法，彼等各係傾向於被感知為清楚分開，此種方法包含之步驟有：

計算上述音訊信號之連續時間區段間的頻譜內容中之差異，其中，該差異係藉由比較連續時間區段間之比響度中的差異來加以計算，其中，比響度係感知響度為頻率和時間之一個函數的計量；以及

在連續時間區段間之頻譜內容中的差異，超過一個臨界值時，識別一個聽覺事件邊界，為此等連續時間區

段間之邊界。

13. 如申請專利範圍第12項之方法，其中，該音訊信號係由一個離散時間序列 $x[n]$ 來表示，其係已在一個取樣頻率 f_s 下取樣自一個音訊源，以及該差異在計算上，係藉由比較橫跨連續時間區段 t 間之頻帶 b 的比響度 $N[b,t]$ 中之差異。

14. 如申請專利範圍第13項之方法，其中，上述音訊信號之連續時間區段間的頻譜內容中之差異，在計算上係依據：

$$D[t] = \sum_b |N_{NORM}[b,t] - N_{NORM}[b,t-1]|$$

其中

$$N_{NORM}[b,t] = \frac{N[b,t]}{\max_b \{N[b,t]\}}。$$

15. 如申請專利範圍第13項之方法，其中，上述音訊信號之連續時間區段間的頻譜內容中之差異，在計算上係依據：

$$D[t] = \sum_b |N_{NORM}[b,t] - N_{NORM}[b,t-1]|$$

其中

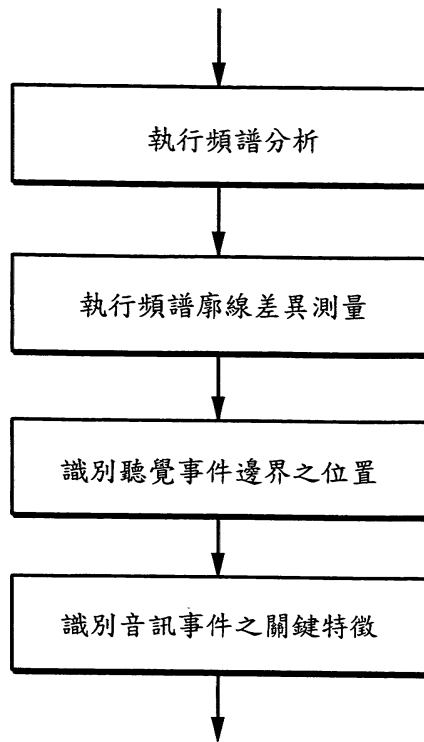
$$N_{NORM}[b,t] = \frac{N[b,t]}{\text{avg}_b \{N[b,t]\}}。$$

16. 一種可被配接來執行申請專利範圍第12至15項任一項之方法的裝置。

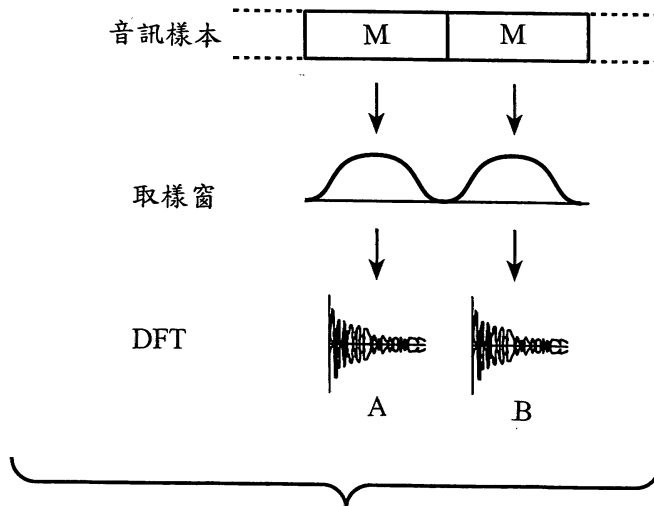
17. 一種電腦程式，其係儲存在一個電腦可讀取式媒體上面，而可使一個電腦控制如申請專利範圍第16項之裝

置。

18. 一種電腦程式，其係儲存在一個電腦可讀取式媒體上面，而可使一個電腦控制如申請專利範圍第12至15項任一項之方法。

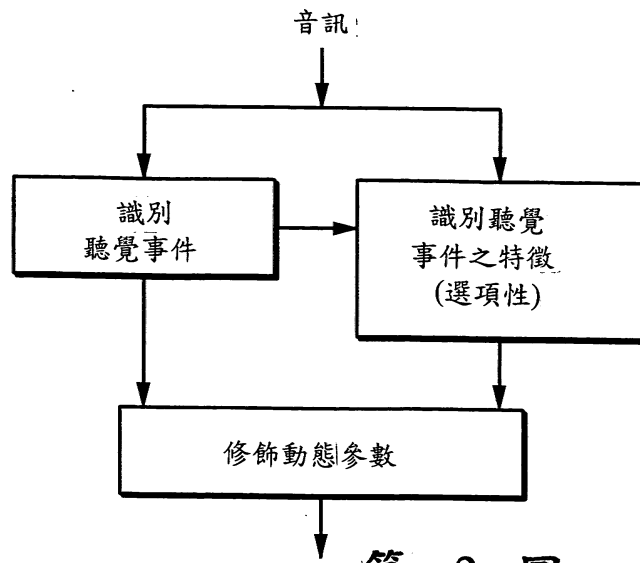


第 1 圖

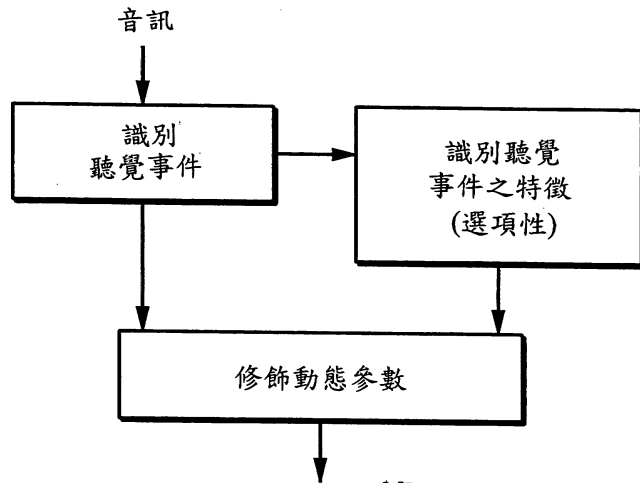


第 2 圖

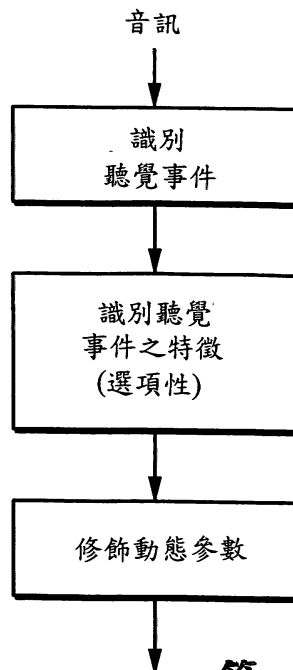
2 / 6



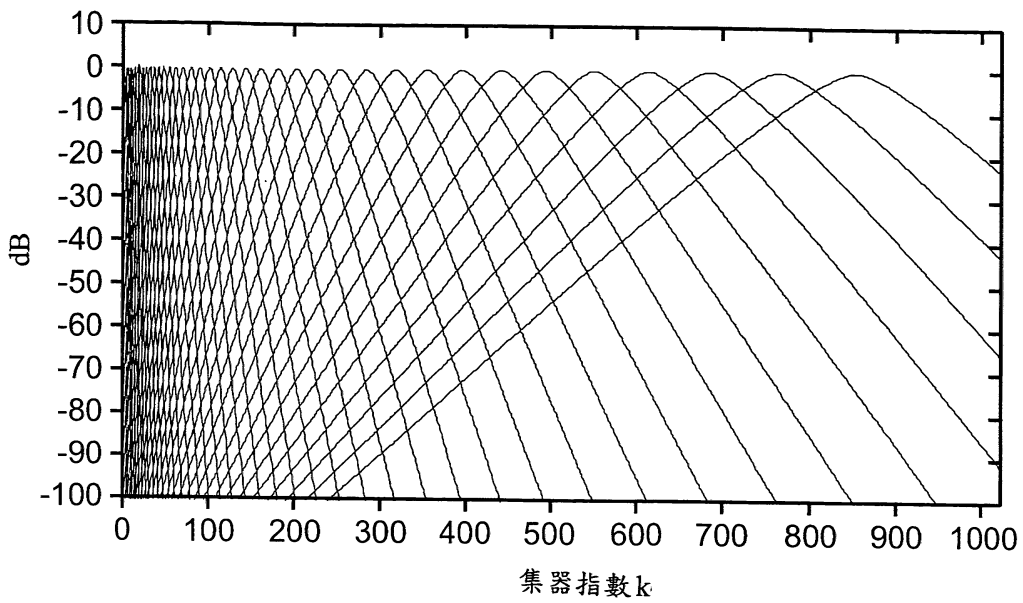
第 3 圖



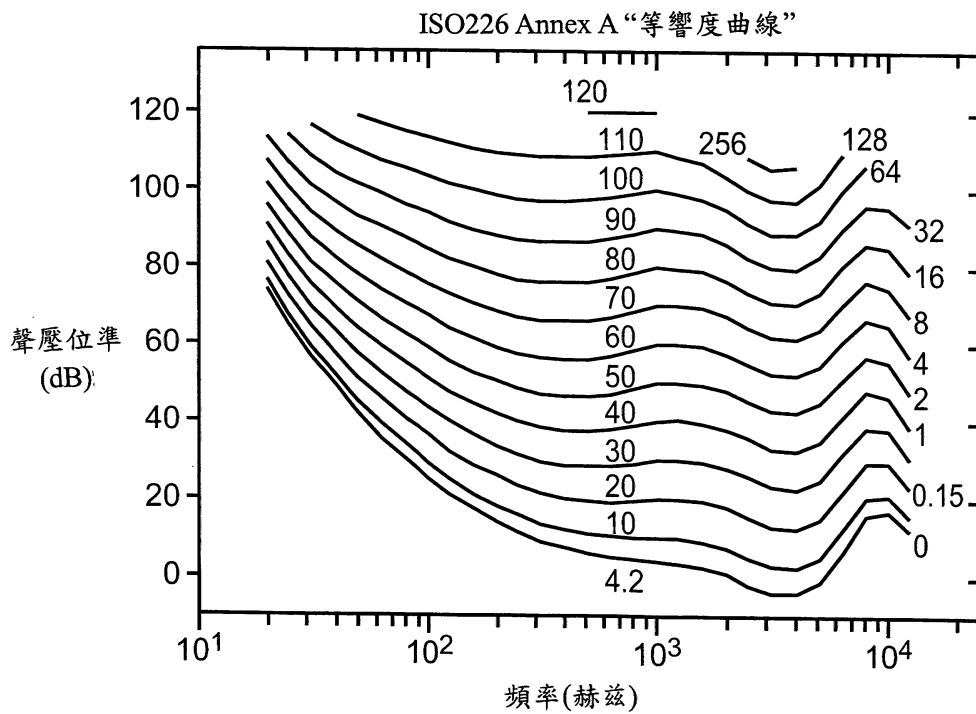
第 4 圖



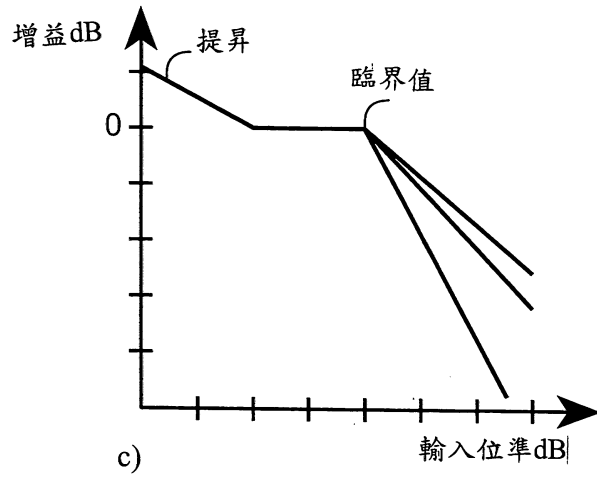
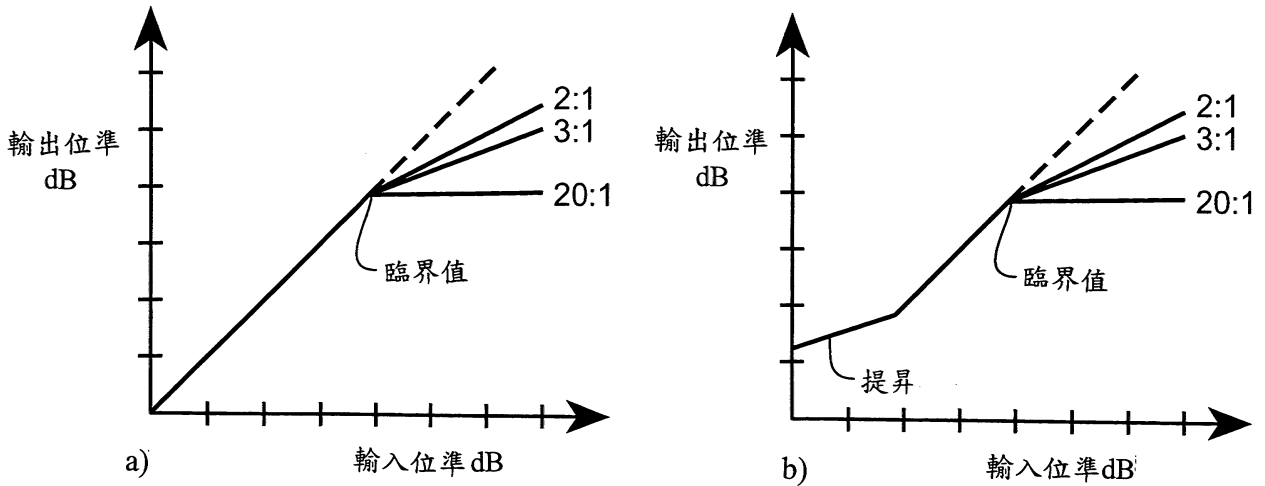
第 5 圖



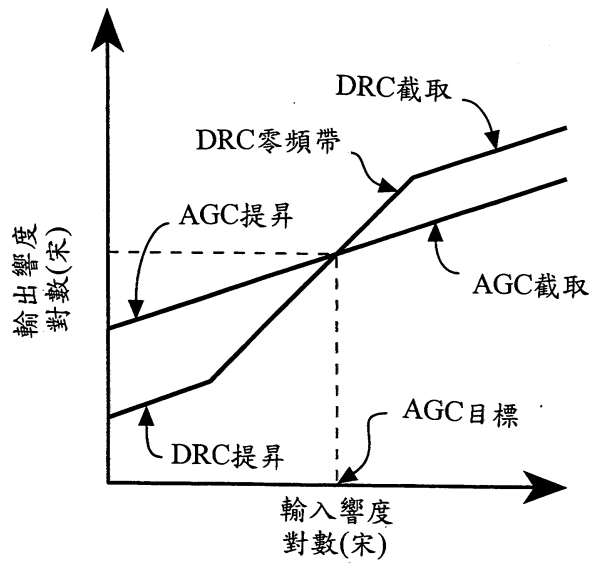
第 6 圖



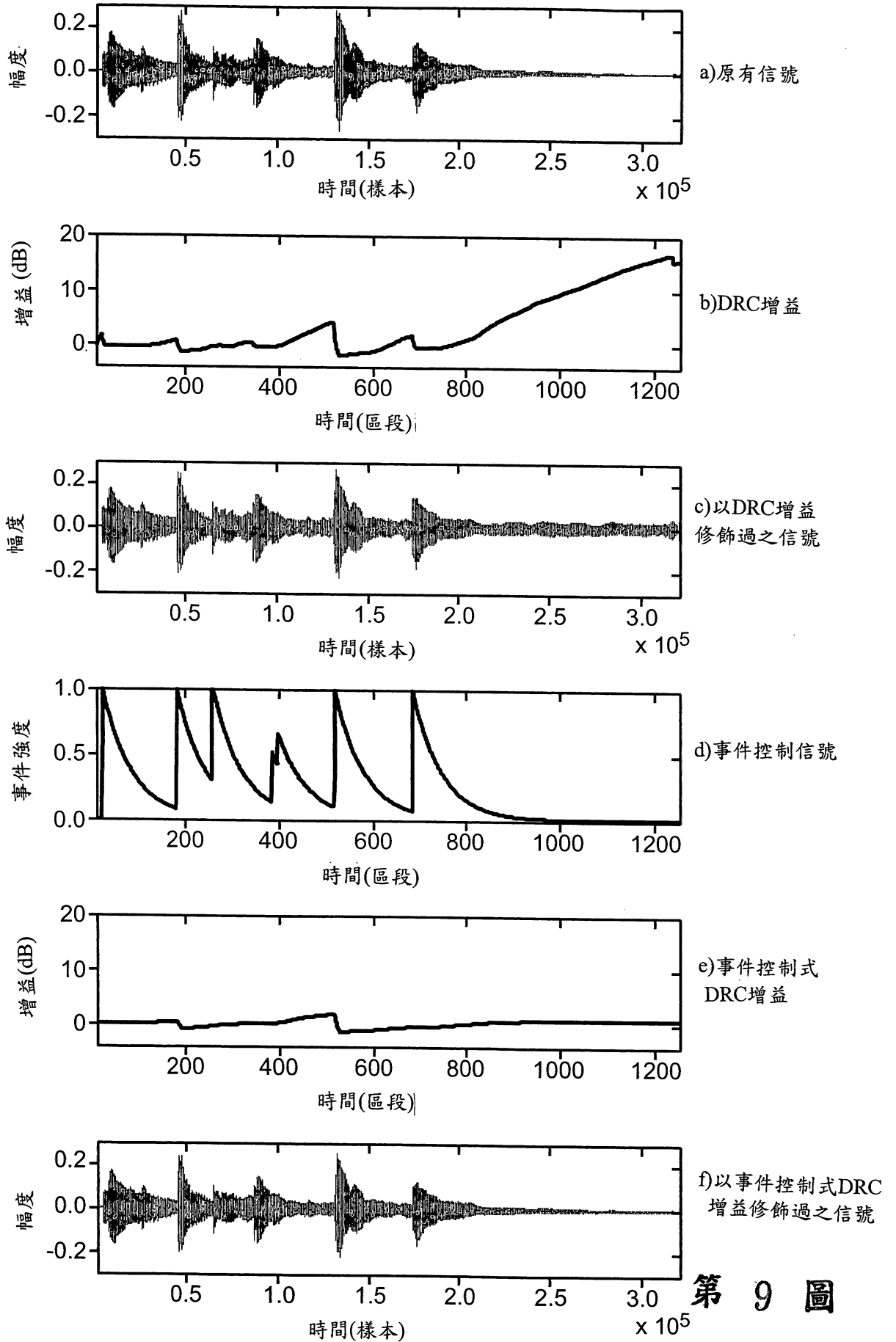
第 7 圖



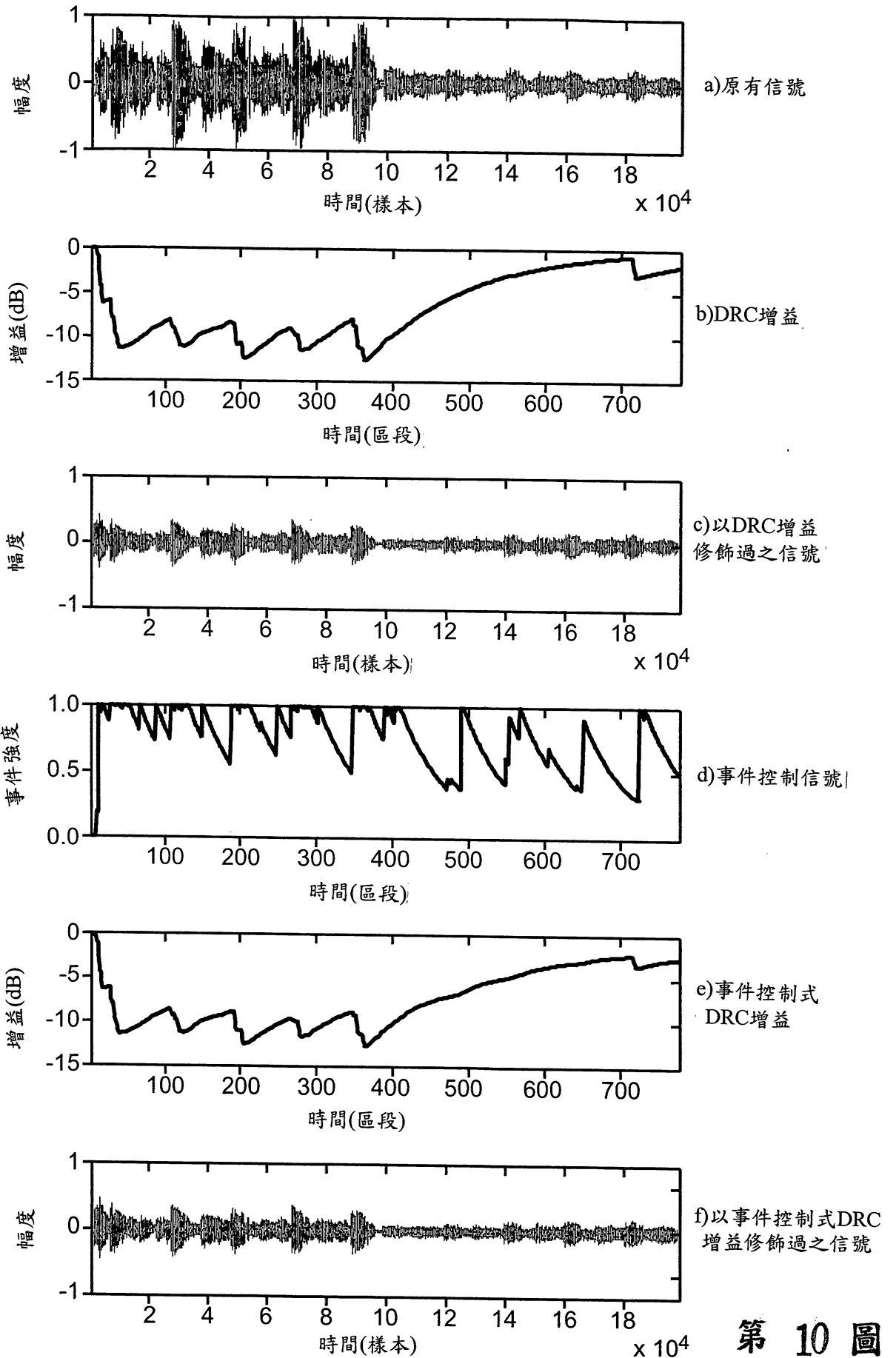
第 8 圖



第 11 圖



6 / 6



第 10 圖

七、指定代表圖：

(一)本案指定代表圖為：第 (3) 圖。

(二)本代表圖之元件符號簡單說明：

(無)

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：