



(19)中華民國智慧財產局

(12)發明說明書公告本 (11)證書號數：TW I489451 B

(45)公告日：中華民國 104 (2015) 年 06 月 21 日

(21)申請案號：101147277

(22)申請日：中華民國 101 (2012) 年 12 月 13 日

(51)Int. Cl. : G10L25/63 (2013.01)

(71)申請人：國立交通大學(中華民國) NATIONAL CHIAO TUNG UNIVERSITY (TW)

新竹市大學路 1001 號

(72)發明人：宋開泰 SONG, KAI TAI (TW)；史 威德 CARLOS, CERVANTES (CO)

(74)代理人：詹銘文；葉璟宗

(56)參考文獻：

US 2009/0164131A1

WO 2012/019827A1

審查人員：机亮燁

申請專利範圍項數：10 項 圖式數：5 共 21 頁

(54)名稱

基於語音情感辨識之音樂播放系統及其方法

MUSIC PLAYING SYSTEM AND METHOD BASED ON SPEECH EMOTION RECOGNITION

(57)摘要

一種音樂播放系統及其方法，適用在基於語音情感辨識以選擇曲目播放，此方法包括下述步驟：利用第一資料庫儲存多首歌曲及該些歌曲映射在情緒座標圖上的歌曲情緒座標，利用第二資料庫儲存情緒辨識參數；分析所接收的語音資料並透過第二資料庫比對，獲取此語音資料映射在情緒座標圖之上的當前情緒座標。接著接收目標情緒座標之設定，在當前情緒座標及目標情緒座標相連的喚醒直線上，找尋最接近喚醒直線的至少一特定歌曲情緒座標，並依序自動播放該些歌曲。

A music playing system and method are provided for playing music based on speech emotion recognition. The method includes the following steps: using a first database to store song emotion coordinates of the songs mapping to a emotion coordinate graph; using a second database to store emotion recognition parameters; analyzing the voice data and get the current emotion coordinate of the voice data mapping to the emotion coordinate graph through the second database; then by receiving a target emotional coordinate settings, and setting up a wake-up line connected between the current emotion coordinate and the target emotional coordinate, finding at least one particular song emotion coordinate closest to the wake-up line and playing the song(s) one by one accordingly.

S205~S260 . . . 音樂播放方法流程步驟

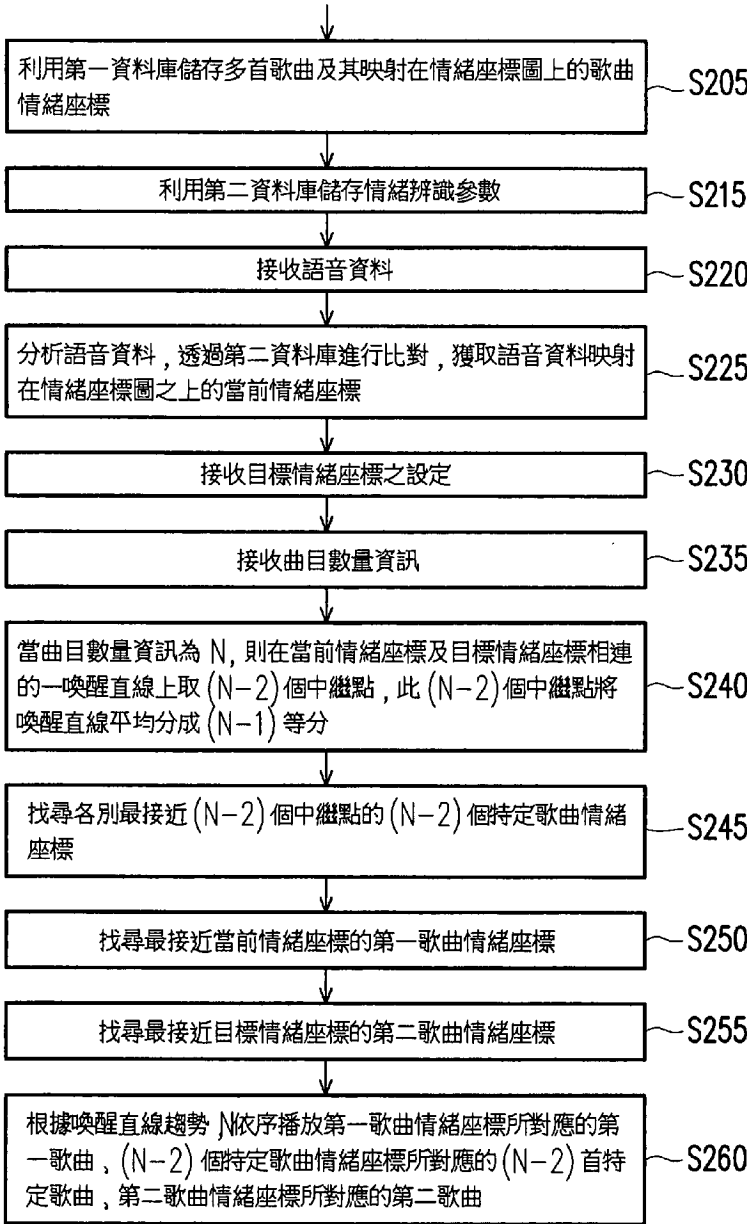


圖 2

公告本

發明專利說明書

(本說明書格式、順序，請勿任意更動，※記號部分請勿填寫)

※申請案號：101147277

※申請日：101.12.13

※IPC 分類：G10L 25/63 (2013.01)

一、發明名稱：

基於語音情感辨識之音樂播放系統及其方法

MUSIC PLAYING SYSTEM AND METHOD BASED
ON SPEECH EMOTION RECOGNITION

二、中文發明摘要：

一種音樂播放系統及其方法，適用在基於語音情感辨識以選擇曲目播放，此方法包括下述步驟：利用第一資料庫儲存多首歌曲及該些歌曲映射在情緒座標圖上的歌曲情緒座標，利用第二資料庫儲存情緒辨識參數；分析所接收的語音資料並透過第二資料庫比對，獲取此語音資料映射在情緒座標圖之上的當前情緒座標。接著接收目標情緒座標之設定，在當前情緒座標及目標情緒座標相連的喚醒直線上，找尋最接近喚醒直線的至少一特定歌曲情緒座標，並依序自動播放該些歌曲。

三、英文發明摘要：

A music playing system and method are provided for playing music based on speech emotion recognition. The

method includes the following steps: using a first database to store song emotion coordinates of the songs mapping to a emotion coordinate graph; using a second database to store emotion recognition parameters; analyzing the voice data and get the current emotion coordinate of the voice data mapping to the emotion coordinate graph through the second database; then by receiving a target emotional coordinate settings, and setting up a wake-up line connected between the current emotion coordinate and the target emotional coordinate, finding at least one particular song emotion coordinate closest to the wake-up line and playing the song(s) one by one accordingly.

四、指定代表圖：

(一) 本案之指定代表圖：圖 2

(二) 本代表圖之元件符號簡單說明：

S205~S260：音樂播放方法流程步驟

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無

method includes the following steps: using a first database to store song emotion coordinates of the songs mapping to a emotion coordinate graph; using a second database to store emotion recognition parameters; analyzing the voice data and get the current emotion coordinate of the voice data mapping to the emotion coordinate graph through the second database; then by receiving a target emotional coordinate settings, and setting up a wake-up line connected between the current emotion coordinate and the target emotional coordinate, finding at least one particular song emotion coordinate closest to the wake-up line and playing the song(s) one by one accordingly.

四、指定代表圖：

(一) 本案之指定代表圖：圖 2

(二) 本代表圖之元件符號簡單說明：

S205~S260：音樂播放方法流程步驟

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無

六、發明說明：

【發明所屬之技術領域】

本發明是有關於一種音樂播放系統及方法，且特別是有關於一種基於語音情感辨識之音樂播放方法與系統。

【先前技術】

就現有技術而言，智慧型、互動型的音樂播放系統尚不多見，且大多需要使用者手動地輸入辨識資料。例如，美國第 7764311 號專利之技術特徵，利用一台攝影機，以人臉圖像特徵為基礎，來辨識面部表情，接著根據檢測到的使用者情緒，播放媒體(包括視頻、音樂)等內容，此方法需要一個具有高解析度且備有良好計算機處理能力的攝像頭，耗費成本較高。

此外，在美國專利申請案第 20090182736 號也提出了歌曲的搜索方法，此為利用已建立之歌曲與情感資訊之對應資料庫，藉由使用者手動輸入目前的情緒資料，擷取適合的曲目，以供使用者聆聽。但是，由於使用者當前的情緒資料，需由使用者手動鍵入，換言之，此乃是使用者主觀決定目前想聽的歌曲，而非外部利用客觀機制判定。

【發明內容】

本發明提出一種音樂播放系統，適用在基於語音情感辨識以選擇曲目播放，此音樂播放系統包括：一第一資料庫、一第二資料庫、一收音裝置及一控制裝置。其中，第

一資料庫儲存多首歌曲及該些歌曲映射在一情緒座標圖上的歌曲情緒座標，第二資料庫則是儲存情緒辨識參數。收音裝置用於接收一語音資料，控制裝置連接至第一資料庫、第二資料庫及收音裝置。接著，控制裝置分析語音資料，透過第二資料庫進行比對，獲取語音資料映射在情緒座標圖之上的一當前情緒座標，接著接收一目標情緒座標之設定，在當前情緒座標及目標情緒座標相連的一喚醒直線上，找尋第一資料庫中最接近喚醒直線的至少一特定歌曲情緒座標，進而播放所選定之歌曲。

在本發明之一實施例中，上述之控制裝置切割此語音資料為多個子語音資料，接著分析此多個子語音資料的頻率，根據在此語音資料中該多個子語音資料的頻率組合，透過第二資料庫進行比對，獲取對應於此頻率組合的當前情緒座標。

在本發明之一實施例中，更包括控制裝置接收一所欲播放之曲目數量資訊，當曲目數量資訊為 N ，其中， N 為正整數，則控制裝置在喚醒直線上取 $(N-2)$ 個中繼點，此 $(N-2)$ 個中繼點將喚醒直線平均分成 $(N-1)$ 等分，控制裝置找尋各別最接近該 $(N-2)$ 個中繼點的 $(N-2)$ 個特定歌曲情緒座標，並找尋第一資料庫中最接近當前情緒座標的一第一歌曲情緒座標及最接近目標情緒座標的一第二歌曲情緒座標。

在本發明之一實施例中，更包括一播放裝置連接至控制裝置，根據此喚醒直線趨勢，播放裝置依序播放第一歌

曲情緒座標所對應的第一歌曲、該(N-2)個特定歌曲情緒座標所對應的(N-2)首特定歌曲、第二歌曲情緒座標所對應的第二歌曲。

在本發明之一實施例中，上述之情緒座標圖為二維座標圖，其中，X 軸為情緒正負向(Valence)參數，Y 軸為情緒強弱向(Arousal)參數。

本發明提出一種音樂播放方法，適用在基於語音情感辨識以選擇曲目播放，包括下述步驟：利用第一資料庫儲存多首歌曲及該些歌曲映射在一情緒座標圖上的歌曲情緒座標；利用第二資料庫儲存情緒辨識參數；接收語音資料；分析語音資料，透過第二資料庫進行比對，獲取語音資料映射在情緒座標圖之上的一當前情緒座標；接著，接收一目標情緒座標之設定；在當前情緒座標及目標情緒座標相連的一喚醒直線上，找尋第一資料庫中最接近喚醒直線的至少一特定歌曲情緒座標。

基於上述，本發明提出一種音樂播放系統及其方法。此方法允許音樂播放系統自動從歌曲資料庫中選取音樂，這些被選取之歌曲也以二維 Valence - Arousal 數值呈現。此外，本發明提出一個喚醒策略，如果使用者聲音被偵測為負面情緒，即會自動選取並播放多樣化音樂類型，用於提升使用者情感進而達到更平靜及快樂的境界。

為讓本發明之上述特徵和優點能更明顯易懂，下文特舉實施例，並配合所附圖式作詳細說明如下。

【實施方式】

圖 1 是依據本發明實施例之一種音樂播放系統的功能方塊圖，音樂播放系統 100 是適用在基於使用者語音情感之辨識，用以選擇曲目播放的系統。音樂播放系統 100 包括：控制裝置 110、第一資料庫 120、第二資料庫 125、收音裝置 130 及播放裝置 140，資料庫 120、收音裝置 130 及播放裝置 140 均連接至控制裝置 110。本發明所提出的音樂播放系統 100 可以存在於一智慧型手機、平板電腦、桌上型電腦、筆記型電腦等電子裝置，其中，控制裝置 110 可為硬體及/或軟體所實現的功能模塊，其中，硬體可包括中央處理器、晶片組、微處理器等具有運算功能的硬體設備或上述硬體設備的組合，而軟體則可以是作業系統、驅動程式等等，收音裝置 130 可以是一麥克風，播放裝置 140 可為一般的音樂播放器。而第一資料庫 120、第二資料庫 125 可以是存於電子裝置的記憶體，亦可以是儲存於網路雲端之上。

圖 2 是依據本發明實施例之一種音樂播放方法的流程圖，請同時參照圖 1 及圖 2。首先，音樂播放系統 100 利用第一資料庫 120 儲存多首歌曲及該些歌曲映射在情緒座標圖上的歌曲情緒座標(步驟 S205)，並利用第二資料庫 125 儲存情緒辨識參數(步驟 S215)。而上述所指的情緒座標圖，其為根據心理學家泰耶(Thayer)提出的二維情緒平面所定義，請見圖 3A~圖 3B。

圖 3A 即為心理學上的二維情緒平面(Two-Dimension Emotional Plane)示意圖。心理學家泰耶提出了如圖 1 的情緒平面，圖 1 的 X 軸為情緒正負向(Valence)，往右為正向(Positive)，往左為負向(Negative)。Y 軸為情緒強弱向(Arousal)，往上為強(High)，往下為弱(Low)。在圖 3A 上也定義了不同種類的情緒，例如：興奮(Excited)、快樂(Happy)、愉悅(Pleased)、輕鬆(Relaxed)、寧靜(Peaceful)、冷靜(Calm)、困倦(Sleepy)、無聊(Bored)、悲傷(Sad)、緊張(Nervous)、生氣(Angry)及苦惱(Annoying)等。

圖 3B 是依據本發明實施例所提出之一種情緒座標圖，基於圖 3A 之理論基礎，此情緒座標圖在 X 軸為情緒正負向，X 軸上最大值定義為+1，最小值為-1。Y 軸為情緒強弱向，Y 軸上最大值定義為+1，最小值為-1。在此情緒座標圖中，中央區域定義為中性(Neutral)情緒，情緒座標圖其他區域也定義了驚喜(Surprise)、快樂(Happy)、無聊(Bored)、悲傷(Sad)、生氣(Angry)等情緒。

在步驟 S205~S215，第一資料庫 120 與第二資料庫 125 中的各項資訊建置完成之後，收音裝置 130 接收一使用者的語音資料(步驟 S220)，控制裝置 110 分析此語音資料，透過第二資料庫 125 進行比對，獲取此語音資料映射在上述情緒座標圖之上的一當前情緒座標(步驟 S225)，此步驟包括控制裝置 110 切割語音資料為多個子語音資料，接著分析多個子語音資料的頻率，根據在此語音資料中多個子

語音資料的頻率組合，透過第二資料庫 125 中的資訊，獲取對應於此頻率組合的當前情緒座標。

接著，接收一目標情緒座標之設定(步驟 S230)，使用者可以自行決定想要到達的情緒狀態，對應輸入目標情緒座標，控制裝置 110 則接收使用者對目標情緒座標之設定。接著，使用者根據目前想收聽的曲目數量，對音樂播放系統 100 輸入曲目數量，控制裝置 110 則接收此對曲目數量資訊之設定(步驟 S235)。

圖 4A 是依據本發明實施例之一種計算情緒座標與歌曲情緒座標最小距離之示意圖，圖 4B 是依據本發明實施例之一種基於喚醒直線選取歌曲的示意圖，接下來的步驟將配合圖 4A 及圖 4B 加以說明。在前述的步驟中可得知，在第一資料庫 120 中，每首歌曲有其對應的歌曲情緒座標，如圖 4A 所示的歌曲 A、歌曲 B 及歌曲 C，當系統需判斷哪一個歌曲情緒座標最接近所選定之情緒座標(如圖 4A 中的三角標誌 402)時，本發明將利用計算歐式距離(Euclidean distance)來判定。也就是說，計算出 $d1$ 、 $d2$ 、 $d3$ ，來判斷歌曲 A、歌曲 B、歌曲 C 何者離圖 4A 的所選定之情緒座標 402 最近，最近距離者，即是最接近所選定之情緒座標之歌曲。

在步驟 S235 中，控制裝置 110 已接收到曲目數量資訊。當曲目數量資訊為 N ， N 為正整數，則控制裝置 110 在當前情緒座標及目標情緒座標相連的一喚醒直線上取 $(N-2)$ 個中繼點，此 $(N-2)$ 個中繼點將喚醒直線平均分成 $(N-1)$

等分(步驟 S240)，接著，控制裝置 110 於第一資料庫 120 中，找尋各別最接近此(N-2)個中繼點的(N-2)個特定歌曲情緒座標 (步驟 S245)，找尋最接近當前情緒座標的第一歌曲情緒座標(步驟 S250)，並找尋最接近目標情緒座標的第二歌曲情緒座標(步驟 S255)。最後，根據喚醒直線趨勢，依序播放第一歌曲情緒座標所對應的第一歌曲、(N-2)個特定歌曲情緒座標所對應的(N-2)首特定歌曲、第二歌曲情緒座標所對應的第二歌曲(步驟 S260)。

以圖 4B 為例，曲目數量資訊為 5，則在當前情緒座標(三角標誌 410)及目標情緒座標(方形標誌 420)相連的一喚醒直線上取 3 個中繼點(十字形標誌 412、414、416)，此 3 個中繼點將喚醒直線平均分成 4 等分。接著，找尋各別最接近此 3 個中繼點的 3 個特定歌曲情緒座標(如圖 4B 上的標誌 432、434、436)，找尋最接近當前情緒座標的第一歌曲情緒座標(標誌 430)，並找尋最接近目標情緒座標的第二歌曲情緒座標(標誌 440)。根據喚醒直線趨勢，依序播放第一歌曲情緒座標 430 所對應的第一歌曲、該 3 個特定歌曲情緒座標(432、434、436)所對應的 3 首特定歌曲、第二歌曲情緒座標 440 所對應的第二歌曲。換言之，依序播放 430、432、434、436、440 所對應的歌曲，讓使用者當前的情緒能被喚醒至目標狀態，也就是最接近目標情緒座標 420 的情緒狀態。

圖 5 是依據本發明實施例之一種音樂播放系統的使用者介面示意圖。在使用者介面 500 中，使用者在目標情緒

座標設置中鍵入情緒正負向之值為 0.55、情緒強弱向之值為 0.33 以及曲目數量為 5，本發明提出的音樂播放系統 100 則可選取出目前適合使用者聆聽的 5 首歌曲，在使用者介面 500 的左下方顯示情緒座標圖，此為控制裝置 100 所計算出的結果，另外在使用者介面 500 的右下方顯示每首歌曲的曲目資訊，以及每首歌曲的情緒正負向與情緒強弱向座標。

綜上所述，本發明提出一種音樂播放系統及其方法。互動應用的人機介面設計，利用聲音訊號作為情感辨識基礎，進而創造一個可感知情緒的音樂播放器，以應用於智慧型手機與智慧機器人。由於人類具有複雜的情緒，例如快樂或高興，不能僅藉由某些特定的數據類別來加以區分，因此本發明提出的音樂播放系統可以將簡短的語音投射至二維 Valence - Arousal 情緒座標，使用者任何情感可以以相對應之連續值表示。本發明之方法也允許系統自動從歌曲資料庫中選取音樂，這些被選取歌曲也以二維 Valence - Arousal 數值呈現。此外，我們提出一個喚醒策略，如果使用者聲音被偵測為負面情緒，即會自動選取並播放多樣化音樂類型，用於提升使用者情感進而達到更平靜及快樂的境界，透過使用者情感表達，可以增進人們與科技產品間之互動。本發明之方法可以被應用現於流行且平價嵌入式平台，它提供使用者電子麥克風及觸控式螢幕作為輸入設備。本發明提出之系統可被整合於各種人機介面應用，亦可透過雲端科技建立資料庫及辨識引擎，舉凡陪伴式機器人、汽車音響及通訊設備。其中通訊設備，如

手機，可以根據使用者情緒狀態播放合適音樂，且現今低資源需求與平價之嵌入式系統已普遍應用於各式產品中。

雖然本發明已以實施例揭露如上，然其並非用以限定本發明，任何所屬技術領域中具有通常知識者，在不脫離本發明之精神和範圍內，當可作些許之更動與潤飾，故本發明之保護範圍當視後附之申請專利範圍所界定者為準。

【圖式簡單說明】

圖 1 是依據本發明實施例之一種音樂播放系統的功能方塊圖。

圖 2 是依據本發明實施例之一種音樂播放方法的流程圖。

圖 3A 是一種心理學上的二維情緒平面示意圖。

圖 3B 是依據本發明實施例之一種情緒座標圖。

圖 4A 是依據本發明實施例之一種計算情緒座標與歌曲情緒座標最小距離之示意圖。

圖 4B 是依據本發明實施例之一種基於喚醒直線選取歌曲的示意圖。

圖 5 是依據本發明實施例之一種音樂播放系統的使用者介面示意圖。

【主要元件符號說明】

100：音樂播放系統

110：控制裝置

120：第一資料庫

125：第二資料庫

130：收音裝置

140：播放裝置

402：被選定之情緒座標

410：當前情緒座標

412、414、416：中繼點

420：目標情緒座標

430、432、434、436、440：歌曲情緒座標

500：使用者介面

S205~S250：音樂播放方法的流程步驟

107年10月9日修正頁(本)

七、申請專利範圍：

1. 一種音樂播放系統，適用在基於語音情感辨識以選擇曲目播放，包括：

一第一資料庫，儲存多首歌曲及該些歌曲映射在一情緒座標圖上的歌曲情緒座標；

一第二資料庫，儲存情緒辨識參數；

一收音裝置，接收一語音資料；

一控制裝置，連接至該第一資料庫、該第二資料庫及該收音裝置，分析該語音資料的頻率，透過該第二資料庫進行比對，獲取該語音資料映射在該情緒座標圖之上的一當前情緒座標，接著接收一目標情緒座標之設定，在該當前情緒座標及該目標情緒座標相連的一喚醒直線上，找尋該第一資料庫中最接近該喚醒直線的至少一特定歌曲情緒座標。

2. 如申請專利範圍第 1 項所述之音樂播放系統，其中，該控制裝置切割該語音資料為多個子語音資料，接著分析該多個子語音資料的頻率，根據在該語音資料中該多個子語音資料的頻率組合，透過該第二資料庫進行比對，獲取對應於該頻率組合的該當前情緒座標。

3. 如申請專利範圍第 1 項所述之音樂播放系統，更包括該控制裝置接收一曲目數量資訊，當該曲目數量資訊為 N ，其中， N 為正整數，則該控制裝置在該喚醒直線上取 $(N-2)$ 個中繼點，該 $(N-2)$ 個中繼點將該喚醒直線平均分成 $(N-1)$ 等分，該控制裝置找尋各別最接近該 $(N-2)$ 個中繼點的 $(N-2)$

個特定歌曲情緒座標，並找尋該第一資料庫中最接近該當前情緒座標的一第一歌曲情緒座標及最接近該目標情緒座標的一第二歌曲情緒座標。

4. 如申請專利範圍第3項所述之音樂播放系統，更包括：

一播放裝置，連接至該控制裝置，根據該喚醒直線趨勢，該播放裝置依序播放該第一歌曲情緒座標所對應的第一歌曲、該(N-2)個特定歌曲情緒座標所對應的(N-2)首特定歌曲、該第二歌曲情緒座標所對應的第二歌曲。

5. 如申請專利範圍第1項所述之音樂播放系統，其中該情緒座標圖為二維座標圖，其中，X軸為情緒正負向(Valence)參數，Y軸為情緒強弱向(Arousal)參數。

6. 一種音樂播放方法，適用在基於語音情感辨識以選擇曲目播放，包括：

利用一第一資料庫儲存多首歌曲及該些歌曲映射在一情緒座標圖上的歌曲情緒座標；

利用一第二資料庫儲存情緒辨識參數；

接收一語音資料；

分析該語音資料的頻率，透過該第二資料庫進行比對，獲取該語音資料映射在該情緒座標圖之上的一當前情緒座標；

接收一目標情緒座標之設定；以及

在該當前情緒座標及該目標情緒座標相連的一喚醒直線上，找尋該第一資料庫中最接近該喚醒直線的至少一特定歌曲情緒座標。

7. 如申請專利範圍第 6 項所述之音樂播放方法，其中，在分析該語音資料的頻率，透過該第二資料庫進行比對，獲取該語音資料映射在該情緒座標圖之上的該當前情緒座標之步驟包括：

切割該語音資料為多個子語音資料；

分析該多個子語音資料的頻率；以及

根據在該語音資料中該多個子語音資料的頻率組合，透過該第二資料庫進行比對，獲取對應於該頻率組合的該當前情緒座標。

8. 如申請專利範圍第 6 項所述之音樂播放方法，更包括：

該控制裝置接收一曲目數量資訊；

當該曲目數量資訊為 N ，其中， N 為正整數，則在該喚醒直線上取 $(N-2)$ 個中繼點，該 $(N-2)$ 個中繼點將該喚醒直線平均分成 $(N-1)$ 等分；

找尋各別最接近該 $(N-2)$ 個中繼點的 $(N-2)$ 個特定歌曲情緒座標；

找尋該第一資料庫中最接近該當前情緒座標的一第一歌曲情緒座標；以及

找尋該第一資料庫中最接近該目標情緒座標的一第二歌曲情緒座標。

9. 如申請專利範圍第 8 項所述之音樂播放方法，更包括：

根據該喚醒直線趨勢，依序播放該第一歌曲情緒座標所對應的第一歌曲、該(N-2)個特定歌曲情緒座標所對應的(N-2)首特定歌曲、該第二歌曲情緒座標所對應的第二歌曲。

10. 如申請專利範圍第 6 項所述之音樂播放方法，其中該情緒座標圖為二維座標圖，其中，X 軸為情緒正負向參數，Y 軸為情緒強弱向參數。

八、圖式：

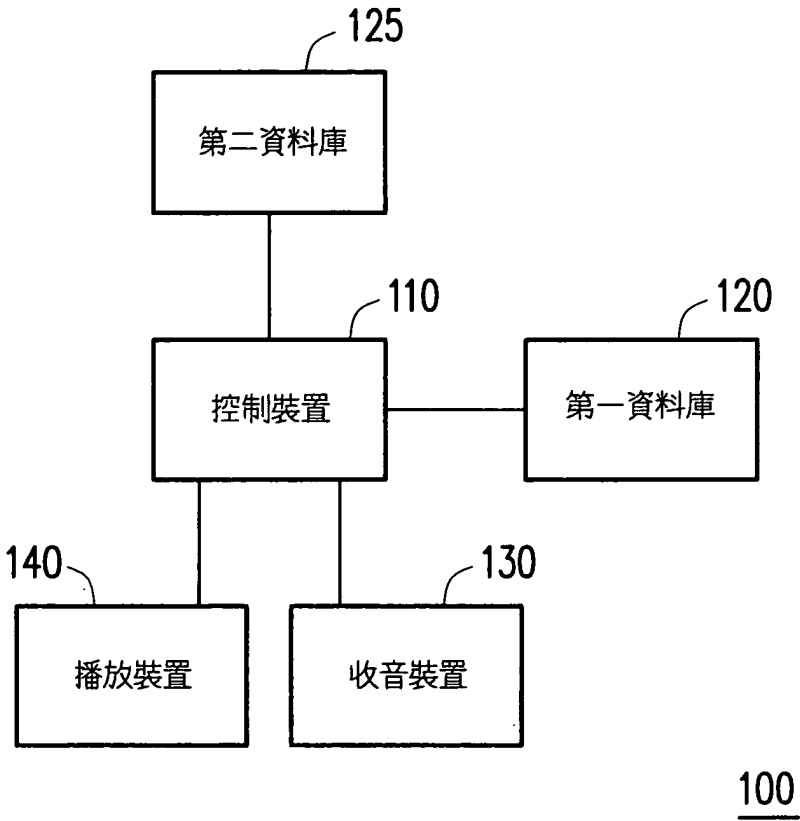


圖 1

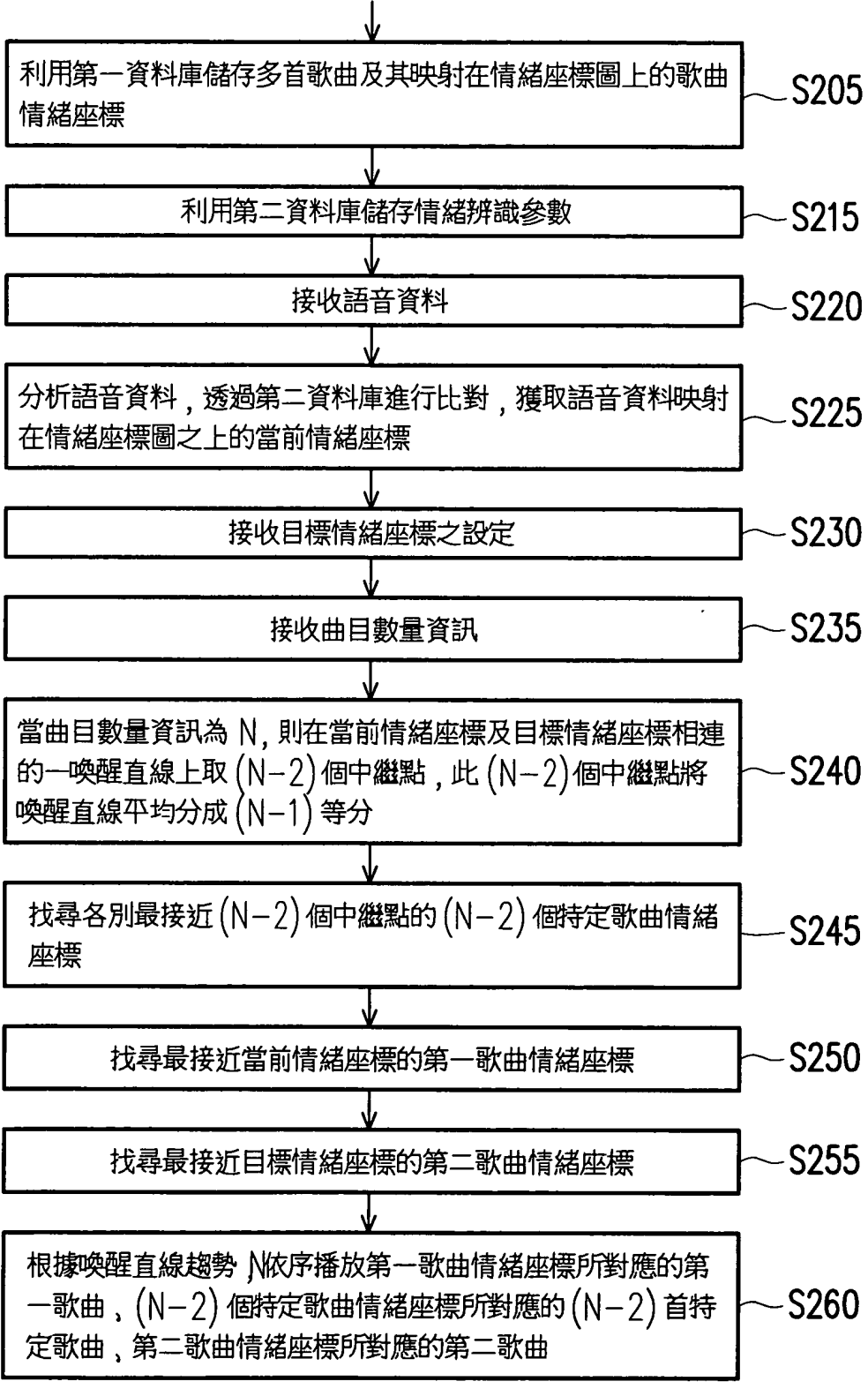


圖 2

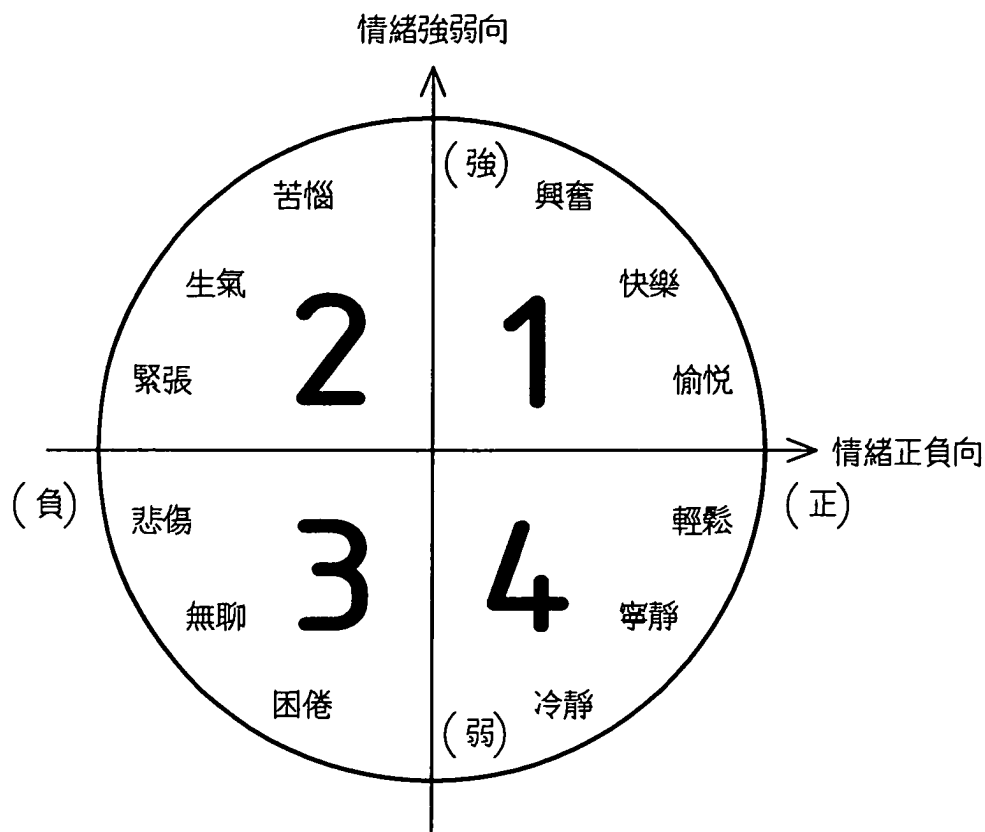


圖 3A

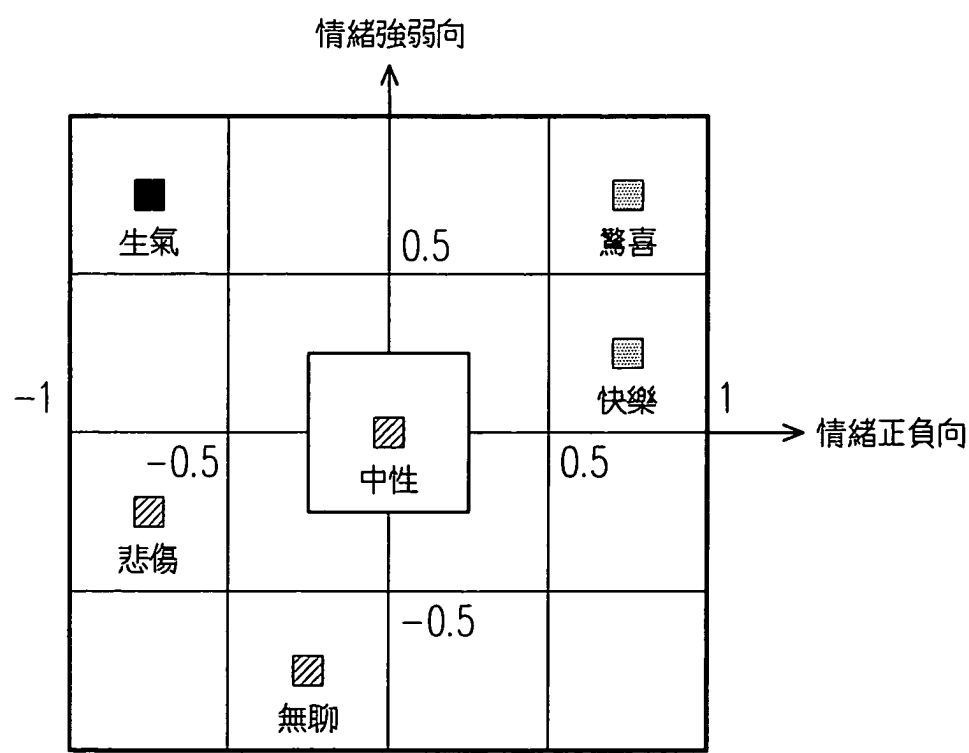


圖 3B

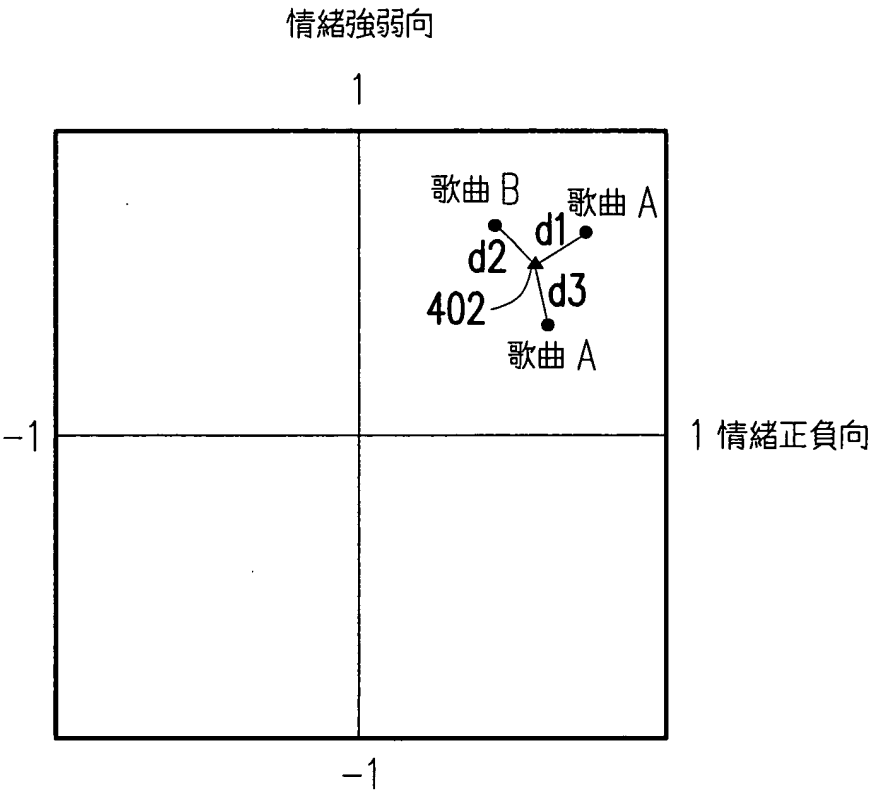


圖 4A

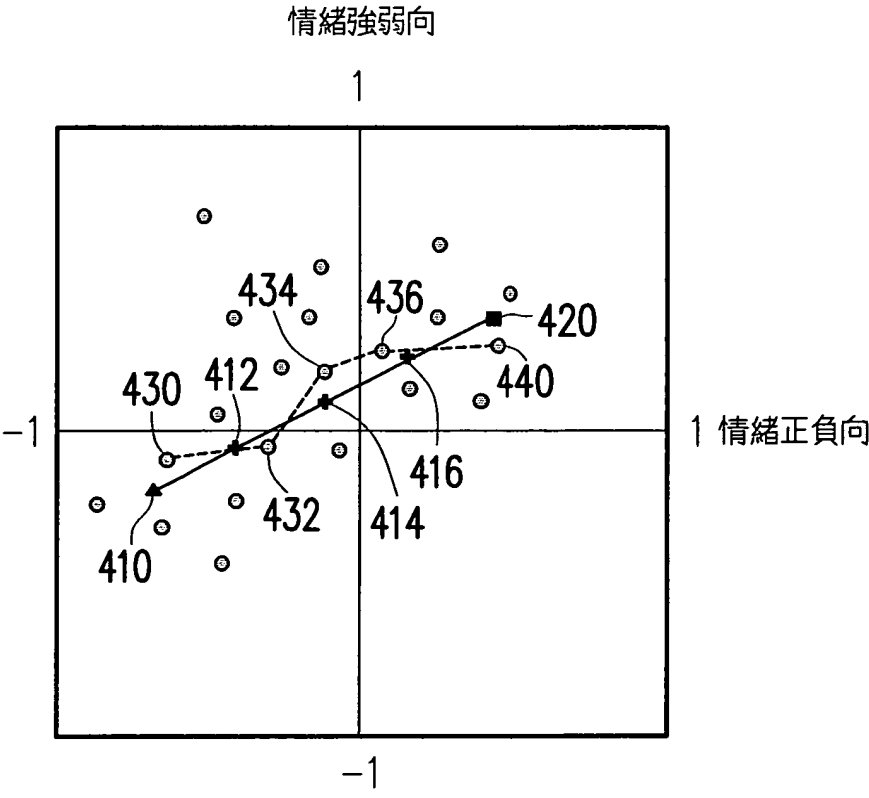


圖 4B

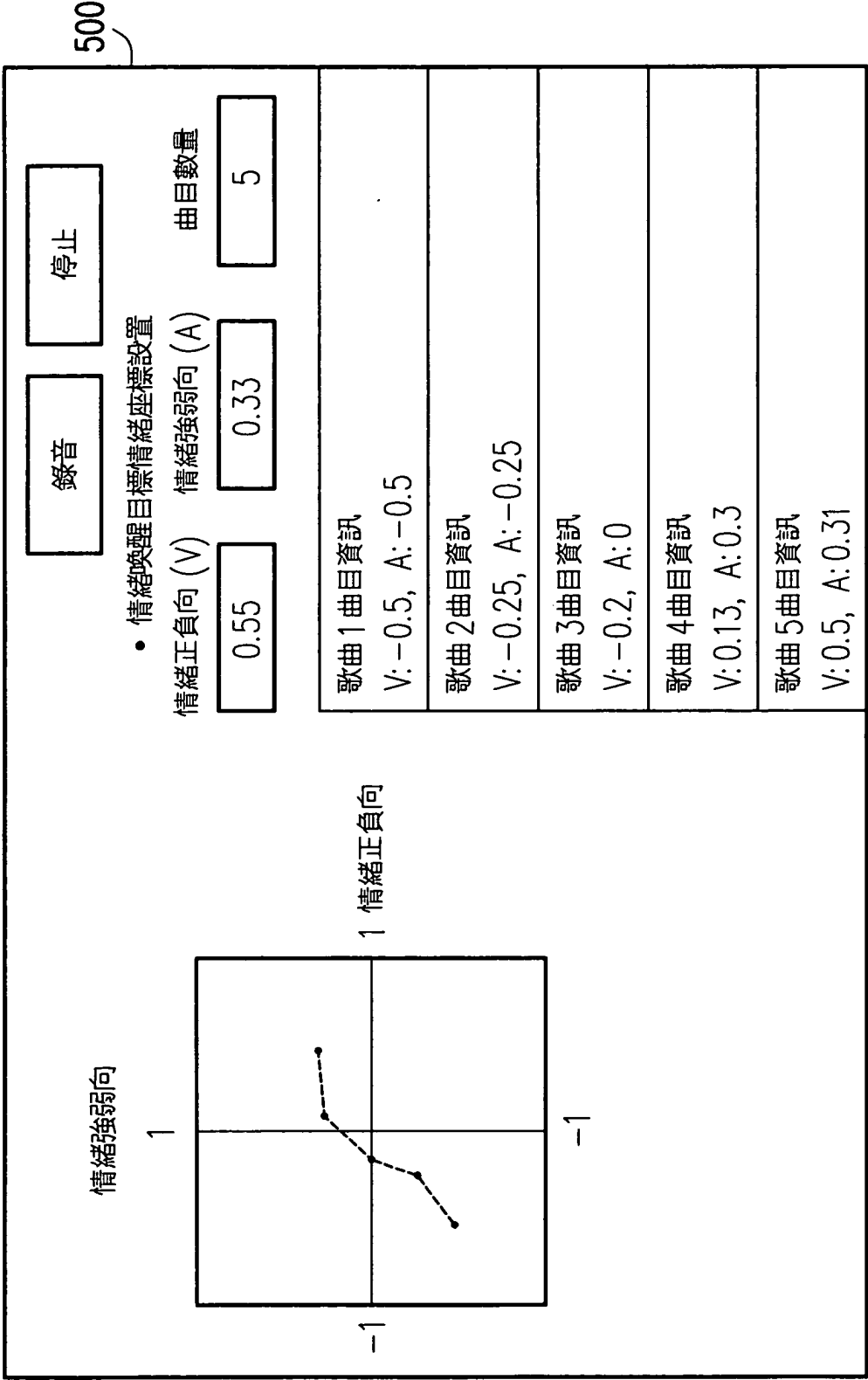


圖5