US 20140172431A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2014/0172431 A1

(43) Pub. Date: Jun. 19, 2014

Song et al.

(54) **MUSIC PLAYING SYSTEM AND MUSIC PLAYING METHOD BASED ON SPEECH EMOTION RECOGNITION**

(71) Applicant: **NATIONAL CHIAO TUNG UNIVERSITY**, Hsinchu City (TW)

(72) Inventors: **Kai-Tai Song**, Hsinchu City (TW); **Carlos Cervantes**, Hsinchu City (TW)

(73) Assignee: **National Chiao Tung University**, Hsinchu City (TW)

(21) Appl. No.: **13/859,760**

(22) Filed: **Apr. 10, 2013**

(30) **Foreign Application Priority Data**

Dec. 13, 2012 (TW) 101147277

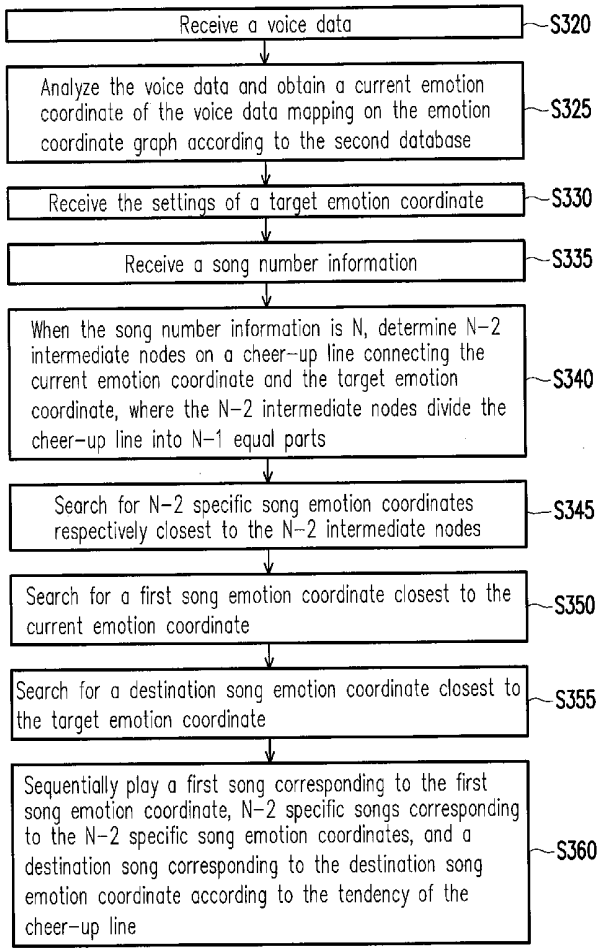
Publication Classification

(51) **Int. Cl.**
G06F 3/16 (2006.01)

(52) **U.S. Cl.**
CPC *G06F 3/167* (2013.01)
USPC **704/275**

(57) **ABSTRACT**

A music playing system and a music playing method suitable for playing music based on speech emotion recognition are provided. The music playing method includes following steps. A plurality of songs and song emotion coordinates of the songs mapping on an emotion coordinate graph are stored in a first database. Emotion recognition parameters are stored in a second database. A voice data is received and analyzed, and a current emotion coordinate of the voice data mapping on the emotion coordinate graph is obtained according to the second database. The setting of a target emotion coordinate is received. At least one specific song emotion coordinate closest to a cheer-up line connecting the current emotion coordinate and the target emotion coordinate is found. Songs corresponding to aforementioned emotion coordinates are sequentially played.



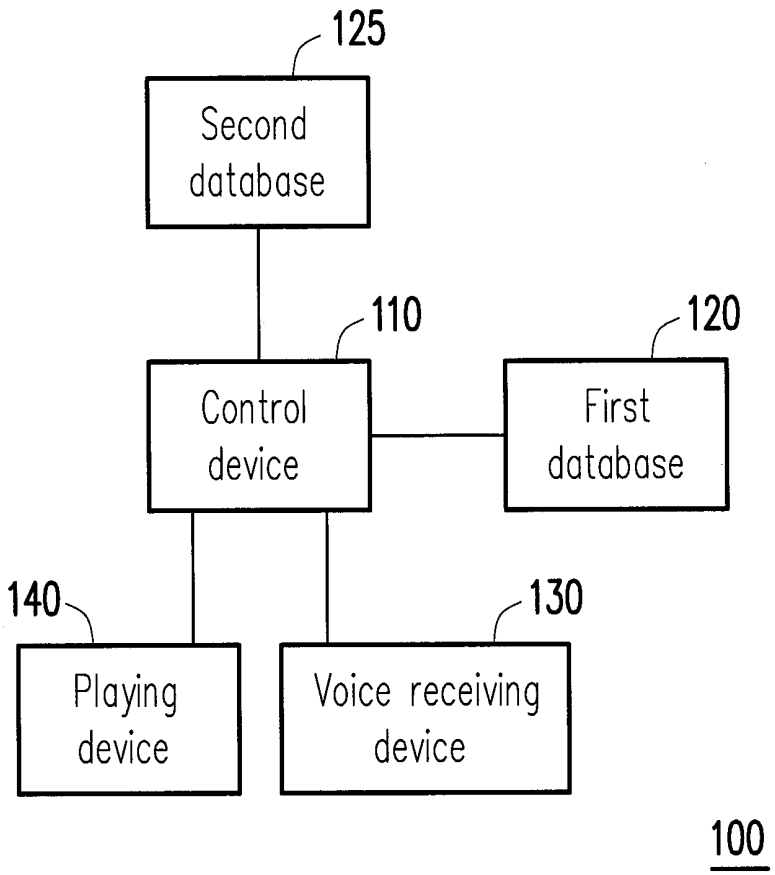

125
Second database
110
120
Control device
First database
140
130
Playing device
Voice receiving device
100

FIG. 1

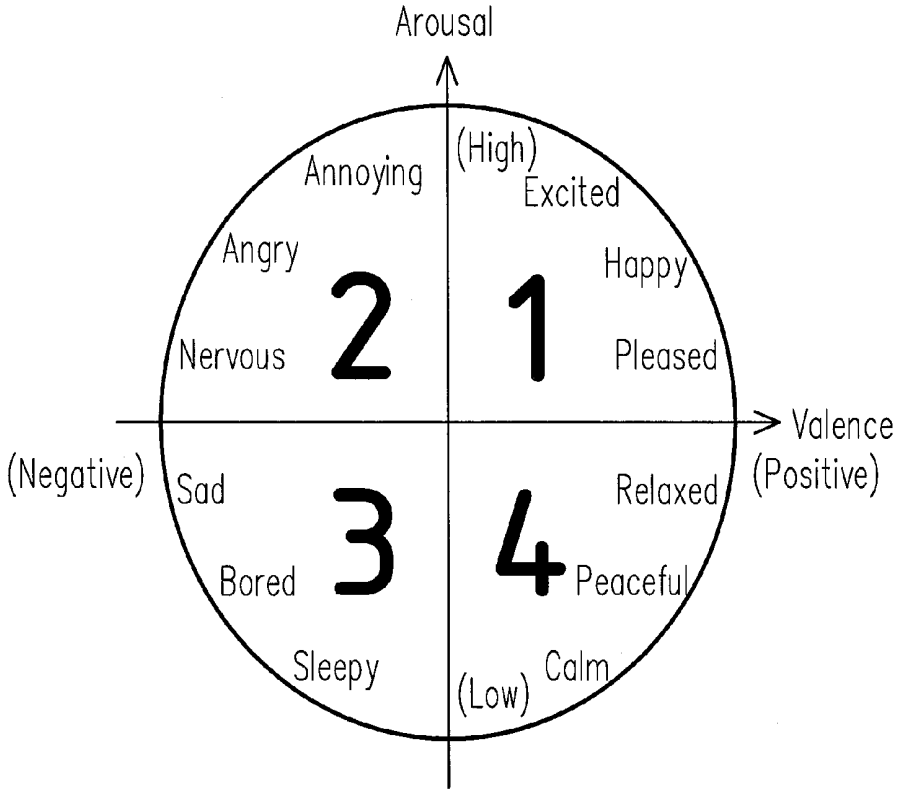

Arousal
(High)
Annoying
Excited
Angry
Happy
2
1
Nervous
Pleased
Valence
(Negative)
(Positive)
Sad
Relaxed
3
4
Bored
Peaceful
Sleepy
Calm
(Low)

FIG. 2A

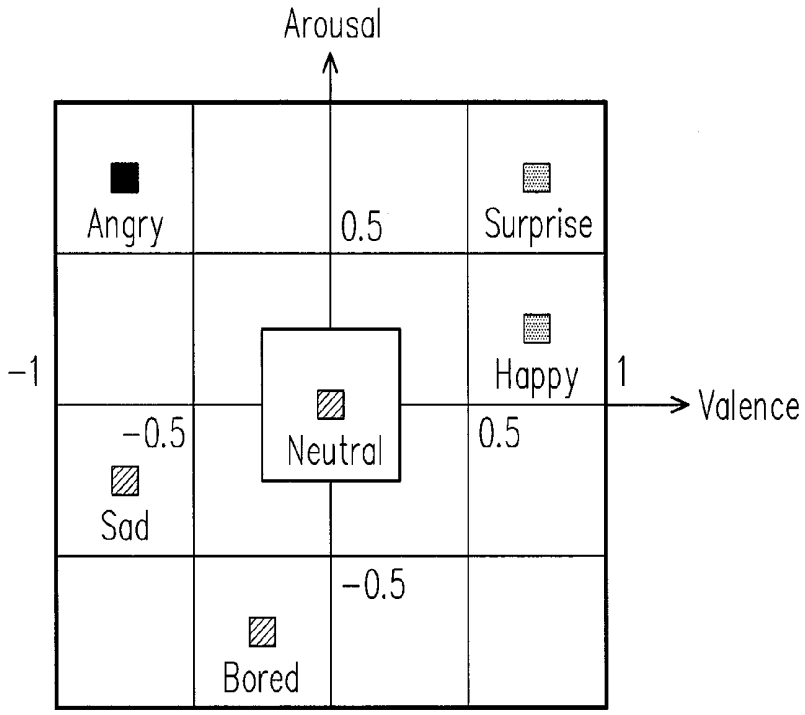

Arousal
Angry
0.5
Surprise
Happy
-1
1
Valence
-0.5
Neutral
0.5
Sad
-0.5
Bored

FIG. 2B

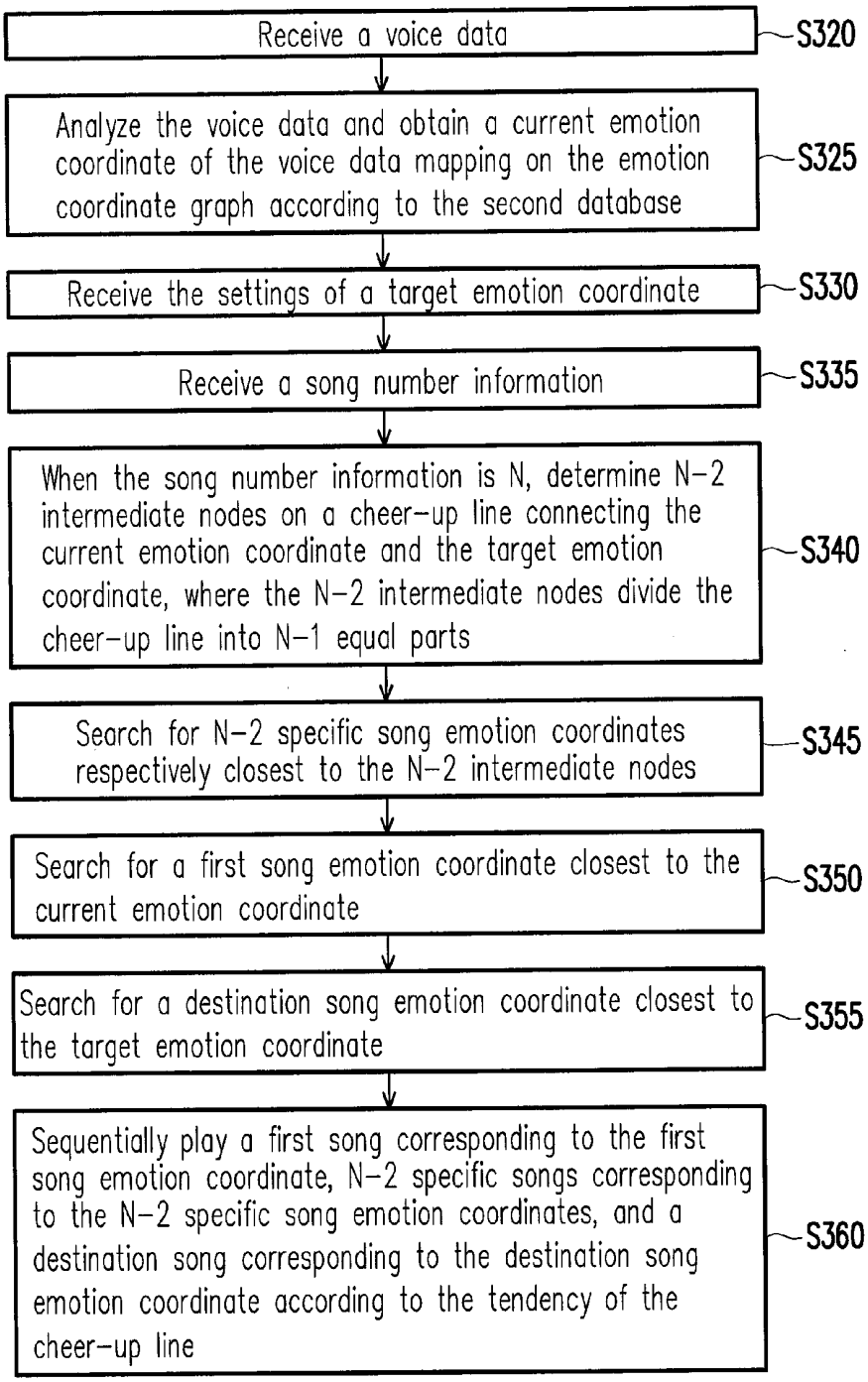

Receive a voice data
S320
Analyze the voice data and obtain a current emotion coordinate of the voice data mapping on the emotion coordinate graph according to the second database
S325
Receive the settings of a target emotion coordinate
S330
Receive a song number information
S335
When the song number information is N, determine N-2 intermediate nodes on a cheer-up line connecting the current emotion coordinate and the target emotion coordinate, where the N-2 intermediate nodes divide the cheer-up line into N-1 equal parts
S340
Search for N-2 specific song emotion coordinates respectively closest to the N-2 intermediate nodes
S345
Search for a first song emotion coordinate closest to the current emotion coordinate
S350
Search for a destination song emotion coordinate closest to the target emotion coordinate
S355
Sequentially play a first song corresponding to the first song emotion coordinate, N-2 specific songs corresponding to the N-2 specific song emotion coordinates, and a destination song corresponding to the destination song emotion coordinate according to the tendency of the cheer-up line
S360

FIG. 3

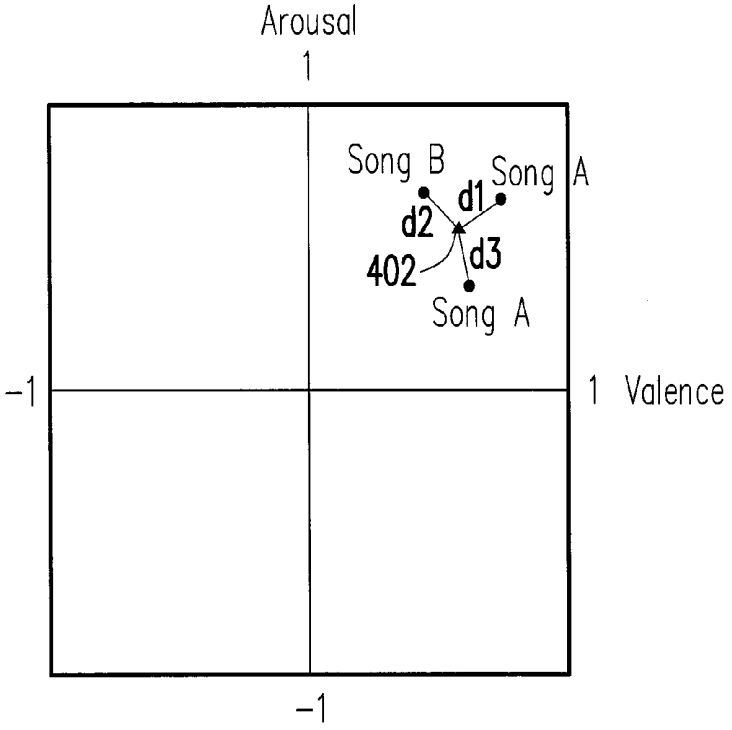

Arousal
1
Song B
Song A
d1
d2
d3
402
Song A
-1
1 Valence
-1

FIG. 4A

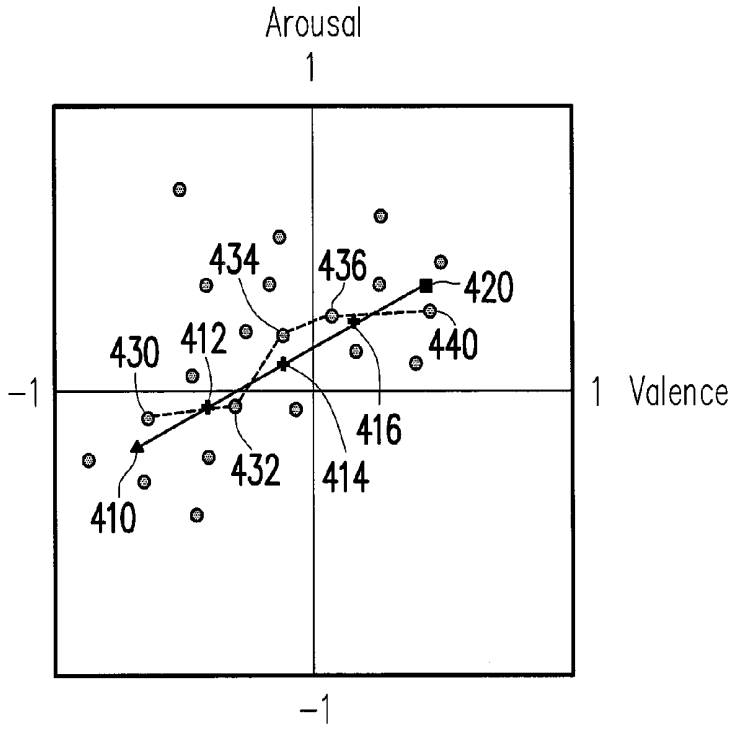

Arousal
1
434
436
420
430
412
440
-1
1 Valence
416
432
414
410
-1

FIG. 4B

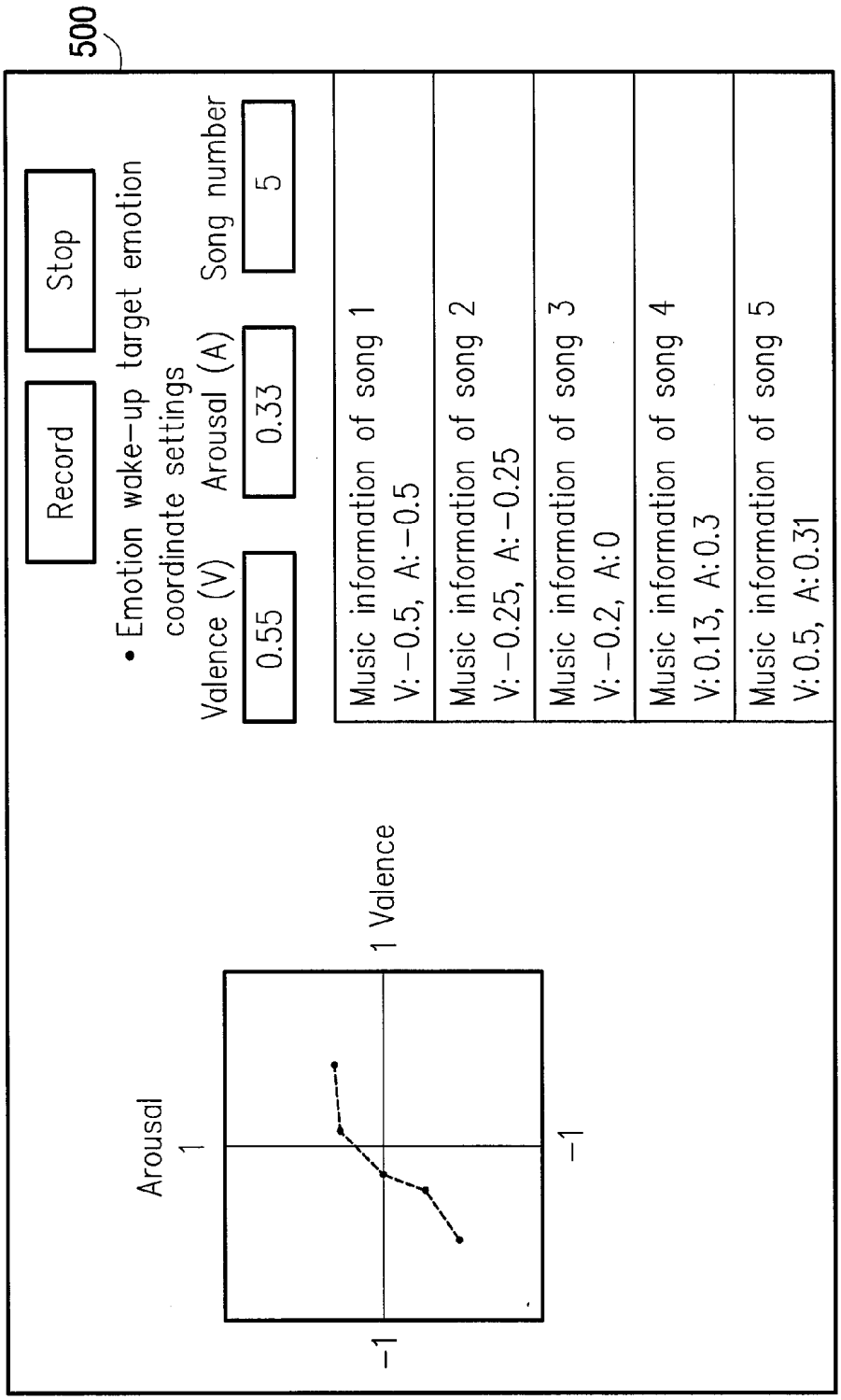

500
Record
Stop
• Emotion wake-up target emotion coordinate settings
Valence (V)
Arousal (A)
Song number
0.55
0.33
5
Music information of song 1
V:−0.5, A:−0.5
Music information of song 2
V:−0.25, A:−0.25
Music information of song 3
V:−0.2, A:0
Music information of song 4
V:0.13, A:0.3
Music information of song 5
V:0.5, A:0.31
Arousal
1
−1
1 Valence
−1

FIG. 5

MUSIC PLAYING SYSTEM AND MUSIC PLAYING METHOD BASED ON SPEECH EMOTION RECOGNITION

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims the priority benefit of Taiwan application serial no. 101147277, filed on Dec. 13, 2012. The entirety of the above-mentioned patent application is hereby incorporated by reference herein and made a part of this specification.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The invention generally relates to a music playing system and a music playing method, and more particularly, to a music playing system and a music playing method based on speech emotion recognition.

[0004] 2. Description of Related Art

[0005] Presently, not many smart and interactive music playing systems are available, and most existing ones require the users to manually input selection data. In U.S. Pat. No. 7,764,311, human facial expressions are recognized based on features of human face images captured by a video camera, and after that, multimedia contents (for example, videos and music) are played according to the detected human emotions. This technique requires a camera of high-resolution and optimal processing capability therefore is quite costly.

[0006] Additionally, a song searching method is disclosed in U.S. Application No. 20090182736, in which appropriate music is selected and played according to a pre-established song-emotion mapping database and current emotion data manually input by a user. However, because the current emotion data of the user needs to be manually input, the song that is selected and played is determined by the user subjectively instead of through an objective mechanism.

SUMMARY OF THE INVENTION

[0007] The invention provides a music playing system suitable for playing music based on speech emotion recognition. The music playing system includes a first database, a second database, a voice receiving device, and a control device. The first database stores a plurality of songs and song emotion coordinates of the songs mapping on an emotion coordinate graph. The second database stores emotion recognition parameters. The voice receiving device receives a voice data. The control device is connected to the first database, the second database, and the voice receiving device. The control device analyzes the voice data and obtains a current emotion coordinate of the voice data mapping on the emotion coordinate graph according to the second database. After that, the control device receives the setting of a target emotion coordinate, searches for at least one specific song emotion coordinate closest to a cheer-up line connecting the current emotion coordinate and the target emotion coordinate in the first database, and plays selected songs sequentially.

[0008] According to an embodiment of the invention, the control device divides the voice data into a plurality of sub voice data, analyzes frequencies of the sub voice data, and obtains the current emotion coordinate corresponding to a frequency combination of the sub voice data in the voice data according to the second database.

[0009] According to an embodiment of the invention, the control device further receives the information of the number of songs to be played in a session. When the song number information is N (where N is a positive integer), the control device determines N−2 intermediate nodes on the cheer-up line, where the N−2 intermediate nodes evenly divide the cheer-up line into N−1 parts. The control device determines N−2 specific song emotion coordinates respectively closest to the N−2 intermediate nodes and searches for a first song emotion coordinate closest to the current emotion coordinate, N−2 specific songs corresponding to the N−2 specific song emotion coordinates, and a destination song emotion coordinate closest to the target emotion coordinate in the first database.

[0010] According to an embodiment of the invention, the music playing system further includes a playing device connected to the control device. The playing device sequentially plays a first song corresponding to the first song emotion coordinate, N−2 specific songs corresponding to the N−2 specific song emotion coordinates, and a destination song corresponding to the destination song emotion coordinate according to a tendency of the cheer-up line.

[0011] According to an embodiment of the invention, the emotion coordinate graph is a two-dimensional coordinate graph, the X-axis of the two-dimensional coordinate graph indicates an emotion valence parameter, and the Y-axis of the two-dimensional coordinate graph indicates an emotion arousal parameter.

[0012] The invention provides a music playing method suitable for playing music based on speech emotion recognition. The music playing method includes following steps. A plurality of songs and song emotion coordinates of the songs mapping on an emotion coordinate graph are stored in a first database. Emotion recognition parameters are stored in a second database. A voice data is received. The voice data is analyzed, and a current emotion coordinate of the voice data mapping on the emotion coordinate graph is obtained according to the second database. The setting of a target emotion coordinate is received. At least one specific song emotion coordinate closest to a cheer-up line connecting the current emotion coordinate and the target emotion coordinate is searched for in the first database.

[0013] As described above, the invention provides a music playing system and a method thereof. The method allows the music playing system to automatically select songs from a song database, wherein the selected songs are presented as two-dimensional valence-arousal values. In addition, the invention provides a cheer-up strategy. If the voice of a user is recognized as carrying a negative emotion, different types of music is automatically selected and played to boost the mood of the user and bring peacefulness and happiness to the user's mind.

[0014] In order to make the aforementioned features and advantages of the invention more comprehensible, embodiments accompanying figures are described in detail below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The accompanying drawings are included to provide a further understanding of the invention, and are incorporated in and constitute a part of this specification. The drawings illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention.

[0016] FIG. **1** is a function block diagram of a music playing system according to an embodiment of the invention.

[0017] FIG. **2**A is a diagram of a psychological two-dimension emotional plane.

[0018] FIG. **2**B illustrates an emotion coordinate graph according to an embodiment of the invention.

[0019] FIG. **3** is a flowchart of a music playing method according to an embodiment of the invention.

[0020] FIG. **4**A is a diagram illustrating how the minimum distance between an emotion coordinate and a song emotion coordinate is calculated according to an embodiment of the invention.

[0021] FIG. **4**B is a diagram illustrating how songs are selected based on a cheer-up line according to an embodiment of the invention.

[0022] FIG. **5** is a diagram of a user interface of a music playing system according to an embodiment of the invention.

DESCRIPTION OF THE EMBODIMENTS

[0023] Reference will now be made in detail to the present preferred embodiments of the invention, examples of which are illustrated in the accompanying drawings. Wherever possible, the same reference numbers are used in the drawings and the description to refer to the same or like parts.

[0024] FIG. **1** is a function block diagram of a music playing system according to an embodiment of the invention. The music playing system **100** is suitable for playing music based on speech emotion recognition. The music playing system **100** includes a control device **110**, a first database **120**, a second database **125**, a voice receiving device **130**, and a playing device **140**. The first database **120**, the voice receiving device **130**, and the playing device **140** are all connected to the control device **110**. The music playing system **100** in the invention can be deployed in a smart phone, a tablet PC, a desktop PC, or a notebook computer. The control device **110** may be a function module implemented in hardware and/or software form. Herein the hardware may be one or a combination of different hardware devices, such as a central processing unit (CPU), a chipset, or a microprocessor, and the software may be an operating system or a firmware program. The voice receiving device **130** may be a microphone. The playing device **140** may be a general music player. The first database **120** and the second database **125** may be stored in a memory of an electronic device or on a network cloud.

[0025] Referring to FIG. **1**, first, the music playing system **100** uses the first database **120** to store a plurality of songs and song emotion coordinates of the songs mapping on an emotion coordinate graph and uses the second database **125** to store emotion recognition parameters. Aforementioned emotion coordinate graph is defined based on the two-dimension emotional plane brought up by the psychologist Thayer, and which can be referred to FIGS. **2**A-**2**B.

[0026] FIG. **2**A is a diagram of a psychological two-dimension emotional plane. The psychologist Thayer brought up the emotional plane illustrated in FIG. **2**A. In FIG. **2**A, the X-axis represents the emotional valence (positive to the right, and negative to the left), and the Y-axis represents the emotional arousal (which grows higher upwards and lower downwards). Different types of emotions, such as excited, happy, pleased, relaxed, peaceful, calm, sleepy, bored, sad, nervous, angry, and annoying, are defined in FIG. **2**A.

[0027] FIG. **2**B illustrates an emotion coordinate graph according to an embodiment of the invention. Based on the definitions illustrated in FIG. **2**A, the X-axis of the emotion coordinate graph indicates the emotional valence (the maximum value on the X-axis is +1, and the minimum value thereon is −1), and the Y-axis thereof indicates the emotional arousal (the maximum value n the Y-axis is +1, and the minimum value thereon is −1). In this emotion coordinate graph, a neutral emotion is defined in the central area, and emotions like surprise, happy, bored, sad, and angry are defined in other areas of the emotion coordinate graph.

[0028] FIG. **3** is a flowchart of a music playing method according to an embodiment of the invention. Referring to both FIG. **1** and FIG. **3**, after various information are established in the first database **120** and the second database **125**, the voice receiving device **130** receives a voice data from a user (step S**320**). The control device **110** analyzes the voice data and obtains a current emotion coordinate of the voice data mapping on the emotion coordinate graph according to the second database **125** (step S**325**). To be specific, in step S**325**, the control device **110** divides the voice data into a plurality of sub voice data, then analyzes the frequencies of the sub voice data, and obtains the current emotion coordinate corresponding to a frequency combination of the sub voice data in the voice data according to information in the second database **125**.

[0029] The user can determine the emotional state he or she wants to approach and input the target emotion coordinate correspondingly (step S**330**). The control device **110** receives the settings of the target emotion coordinate from the user. Next, the user inputs the number of songs he or she wants to listen into the music playing system **100**, and the control device **110** receives this song number information (step S**335**).

[0030] FIG. **4**A is a diagram illustrating how the minimum distance between an emotion coordinate and a song emotion coordinate is calculated according to an embodiment of the invention. FIG. **4**B is a diagram illustrating how songs are selected based on a cheer-up line according to an embodiment of the invention. Following steps will be explained with reference to both FIG. **4**A and FIG. **4**B. As mentioned in foregoing steps, in the first database **120**, each song has a corresponding song emotion coordinate (for example, the songs A, B, and C in FIG. **4**A). In the invention, the song emotion coordinate closest to a selected emotion coordinate (for example, the triangular mark **402** in FIG. **4**A) is determined by calculating a Euclidean distance. Namely, the Euclidean distances d1, d2, and d3 are calculated to determine which one of the songs A, B, and C is closest to the selected emotion coordinate **402** in FIG. **4**A, and the song has the shortest distance is the song closest to the selected emotion coordinate.

[0031] In step S**335**, the control device **110** receives the song number information. When the song number information is N (where N is a positive integer), the control device **110** determines N−2 intermediate nodes on a cheer-up line connecting the current emotion coordinate and the target emotion coordinate. Herein the N−2 intermediate nodes divide the cheer-up line into N−1 equal parts (step S**340**). After that, the control device **110** searches for N−2 specific song emotion coordinates respectively closest to the N−2 intermediate nodes (step S**345**), a first song emotion coordinate closest to the current emotion coordinate (step S**350**), and a destination song emotion coordinate closest to the target emotion coordinate in the first database **120** (step S**355**). Finally, the control device **110** sequentially plays a first song corresponding to the first song emotion coordinate, N−2 specific songs cor-

responding to the N−2 specific song emotion coordinates, and a destination song corresponding to the destination song emotion coordinate according to the tendency of the cheer-up line (step S360).

[0032] Referring to FIG. 4B, assuming that the song number information is 5, three intermediate nodes (the cross marks **412**, **414**, and **416**) are determined on the cheer-up line connecting the current emotion coordinate (the triangular mark **410**) and the target emotion coordinate (the square mark **420**). Herein the three intermediate nodes divide the cheer-up line into four equal parts. Then, three specific song emotion coordinates respectively closest to the three intermediate nodes (the marks **432**, **434**, and **436** in FIG. 4B), the first song emotion coordinate closest to the current emotion coordinate (the mark **430**), and the destination song emotion coordinate closest to the target emotion coordinate (the mark **440**) are searched. The first song corresponding to the first song emotion coordinate **430**, three specific songs corresponding to the three specific song emotion coordinates **432**, **434**, and **436**, and the destination song corresponding to the destination song emotion coordinate **440** are sequentially played according to the tendency of the cheer-up line. In other words, songs corresponding to the song emotion coordinates **430**, **432**, **434**, **436**, and **440** are sequentially played to bring the current emotion of the user to the target state (i.e., the emotional state closest to the target emotion coordinate **420**).

[0033] FIG. **5** is a diagram of a user interface of a music playing system according to an embodiment of the invention. In the user interface **500**, a user inputs an emotional valence value 0.55, an emotional arousal value 0.33, and a song number **5** as the settings of a target emotion coordinate. Accordingly, the music playing system **100** provided by the invention selects five songs most suitable to the user and displays an emotion coordinate graph at the bottom left part of the user interface **500**. These five songs are determined by the control device **110**. Additionally, music information and valence and arousal coordinates of each song are displayed at the bottom right part of the user interface **500**.

[0034] As described above, the invention provides a music playing system and a method thereof, in which an interactive human-machine interface is provided and voice signals are used for emotion recognition, so that a music player capable of emotion sensing is created and can be applied in smart phones and certain companion robots. Because human beings have complicated emotions (for example, happy or glad), these emotions cannot be distinguished according to some specific data types. Thereby, in the music playing system provided by the invention, a short utterance is projected to a two-dimensional valence-arousal coordinate, and any emotion of a user can be represented by a corresponding value. The method in the invention allows the system to automatically select songs from a song database, and the selected songs are also presented with two-dimensional valence-arousal values. In addition, a cheer-up strategy is provided. If a user's voice is determined to carry negative emotion, different types of music is automatically selected and played to boost the mood of the user and bring the user's mind to a peaceful and happy state. Through such emotion expression, the user can interact with the technical product more actively. The method provided by the invention can be applied to today most popular and inexpensive embedded platforms which provide electronic microphones and touch screens as the input devices. The music playing system in the invention may be integrated with various human-machine interface applications (for example, companion robot, car audio systems, and communication equipments), or the databases and recognition engines may be established through the cloud computing technology. Aforementioned communication device (for example, a cell phone) can play suitable music according to the emotional state of a user. Nowadays, embedded systems with low resource requirement and low price have been broadly applied to various products.

[0035] It will be apparent to those skilled in the art that various modifications and variations can be made to the structure of the invention without departing from the scope or spirit of the invention. In view of the foregoing, it is intended that the invention cover modifications and variations of this invention provided they fall within the scope of the following claims and their equivalents.

What is claimed is:

1. A music playing system, suitable for playing music based on speech emotion recognition, the music playing system comprising:

a first database, storing a plurality of songs and song emotion coordinates of the songs mapping on an emotion coordinate graph;

a second database, storing emotion recognition parameters;

a voice receiving device, receiving a voice data;

a control device, connected to the first database, the second database, and the voice receiving device, analyzing the voice data, obtaining a current emotion coordinate of the voice data mapping on the emotion coordinate graph according to the second database, receiving settings of a target emotion coordinate, and searching for at least one specific song emotion coordinate closest to a cheer-up line connecting the current emotion coordinate and the target emotion coordinate in the first database.

2. The music playing system according to claim **1**, wherein the control device divides the voice data into a plurality of sub voice data, analyzes frequencies of the sub voice data, and obtains the current emotion coordinate corresponding to a frequency combination of the sub voice data in the voice data according to the second database.

3. The music playing system according to claim **1**, wherein the control device further receives a song number information, when the song number information is N (wherein N is a positive integer), the control device determines N−2 intermediate nodes on the cheer-up line, the N−2 intermediate nodes evenly divide the cheer-up line into N−1 parts, the control device determines N−2 specific song emotion coordinates respectively closest to the N−2 intermediate nodes and searches for a first song emotion coordinate closest to the current emotion coordinate and a destination song emotion coordinate closest to the target emotion coordinate in the first database.

4. The music playing system according to claim **3** further comprising:

a playing device, connected to the control device, and sequentially playing a first song corresponding to the first song emotion coordinate, N−2 specific songs corresponding to the N−2 specific song emotion coordinates, and a destination song corresponding to the destination song emotion coordinate according to a tendency of the cheer-up line.

5. The music playing system according to claim **1**, wherein the emotion coordinate graph is a two-dimensional coordinate graph, an X-axis of the two-dimensional coordinate

graph indicates an emotion valence parameter, and a Y-axis of the two-dimensional coordinate graph indicates an emotion arousal parameter.

6. A music playing method, suitable for playing music based on speech emotion recognition, the music playing method comprising:

storing a plurality of songs and song emotion coordinates of the songs mapping on an emotion coordinate graph by using a first database;

storing emotion recognition parameters by using a second database;

receiving a voice data;

analyzing the voice data, and obtaining a current emotion coordinate of the voice data mapping on the emotion coordinate graph according to the second database;

receiving settings of a target emotion coordinate; and

searching for at least one specific song emotion coordinate closest to a cheer-up line connecting the current emotion coordinate and the target emotion coordinate in the first database.

7. The music playing method according to claim **6**, wherein the step of analyzing the voice data and obtaining the current emotion coordinate of the voice data mapping on the emotion coordinate graph according to the second database comprises:

dividing the voice data into a plurality of sub voice data;

analyzing frequencies of the sub voice data; and

obtaining the current emotion coordinate corresponding to a frequency combination of the sub voice data in the voice data according to the second database.

8. The music playing method according to claim **6** further comprising:

receiving a song number information by using a control device;

when the song number information is N (wherein N is a positive integer), determining N−2 intermediate nodes on the cheer-up line, wherein the N−2 intermediate nodes evenly divide the cheer-up line into N−1 parts;

determining N−2 specific song emotion coordinates respectively closest to the N−2 intermediate nodes;

searching for a first song emotion coordinate closest to the current emotion coordinate in the first database; and

searching for a destination song emotion coordinate closest to the target emotion coordinate in the first database.

9. The music playing method according to claim **8** further comprising:

sequentially playing a first song corresponding to the first song emotion coordinate, N−2 specific songs corresponding to the N−2 specific song emotion coordinates, and a destination song corresponding to the destination song emotion coordinate according to a tendency of the cheer-up line.

10. The music playing method according to claim **6**, wherein the emotion coordinate graph is a two-dimensional coordinate graph, an X-axis of the two-dimensional coordinate graph indicates an emotion valence parameter, and a Y-axis of the two-dimensional coordinate graph indicates an emotion arousal parameter.

* * * * *