



(11)

**EP 4 579 653 A1**

(12)

**EUROPEAN PATENT APPLICATION**  
published in accordance with Art. 153(4) EPC

(43) Date of publication:  
**02.07.2025 Bulletin 2025/27**

(51) International Patent Classification (IPC):  
**G10L 13/02 (2013.01)**

(21) Application number: **23825595.4**

(52) Cooperative Patent Classification (CPC):  
**G10L 13/02; G10L 13/033; G10L 13/08;  
G10L 25/03; G10L 25/30**

(22) Date of filing: **30.06.2023**

(86) International application number:  
**PCT/CN2023/105292**

(87) International publication number:  
**WO 2024/087727 (02.05.2024 Gazette 2024/18)**

(84) Designated Contracting States:  
**AL AT BE BG CH CY CZ DE DK EE ES FI FR GB  
GR HR HU IE IS IT LI LT LU LV MC ME MK MT NL  
NO PL PT RO RS SE SI SK SM TR**  
Designated Extension States:  
**BA**  
Designated Validation States:  
**KH MA MD TN**

- **LU, Fang**  
Wuhan, Hubei 430050 (CN)
- **ZHOU, Bing**  
Wuhan, Hubei 430050 (CN)
- **LI, Ping**  
Wuhan, Hubei 430050 (CN)
- **TANG, Mazheng**  
Wuhan, Hubei 430050 (CN)
- **YANG, Jin**  
Wuhan, Hubei 430050 (CN)
- **MIAO, Yudong**  
Wuhan, Hubei 430050 (CN)

(30) Priority: **28.10.2022 CN 202211335986**

(71) Applicant: **Voyah Automobile Technology  
Company Ltd.**  
**Hanyang District**  
**Wuhan City, Hubei 430050 (CN)**

(74) Representative: **Valet Patent Services Limited**  
**c/o Caya 83713X**  
**Am Börstig 5**  
**96052 Bamberg (DE)**

(72) Inventors:  
• **ZHANG, Guihai**  
**Wuhan, Hubei 430050 (CN)**

(54) **VOICE DATA PROCESSING METHOD BASED ON IN-VEHICLE VOICE AI, AND RELATED DEVICE**

(57) This article discloses a voice data processing method based on an on-board voice AI. The method includes: acquiring an audio feature and a lyrics feature of a target song; generating a singing voice data of the on-board voice AI for the target song based on the audio feature and the lyrics feature; broadcasting the target song and the singing voice data of the on-board voice AI at the same time; and collecting and broadcasting an audio data of a target user in real time.

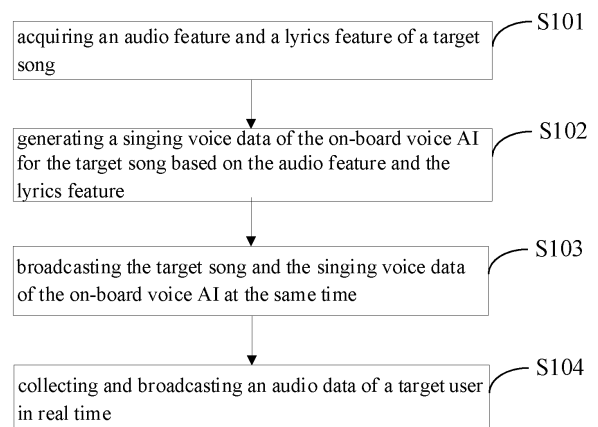


FIG. 1

**EP 4 579 653 A1**

## Description

### CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application claims a priority of the Chinese patent application No. 202211335986.9 filed on October 28, 2022, the entire content of which is incorporated herein by reference.

### TECHNICAL FIELD

**[0002]** The present invention belongs to a field of intelligent voice technology, and in particular relates to a voice data processing method and related device based on an on-board voice AI.

### BACKGROUND

**[0003]** With the improvement of people's living standards and the yearning for a better life, the demand for self-driving entertainment is increasing each day. Currently, inside a vehicle, most on-board cockpits are equipped with a mature karaoke software. A user can use the karaoke software to sing by himself or with other users through a microphone. This type of karaoke software can process and mix a voice from the user to generate a human voice with a reverberation effect, and then mix the human voice with a song accompaniment to produce a sound of singing. However, the user cannot use this type of karaoke software to chorus or duet with an AI assistant. The user can only shift into an original vocal to sing, that is, the user cannot chorus his favorite songs with the AI assistant at the same time. In terms of singing entertainment, an interactivity with the AI assistant is relatively low, resulting in a low competitiveness of intelligent cockpit products equipped with this type of karaoke software.

### SUMMARY

**[0004]** In view of the above problems, the present invention proposes a voice data processing method and related device based on an on-board voice AI, which improves an interactivity between users and AI assistants as well as an entertainment and a product competitiveness of an in-vehicle intelligent cockpit.

**[0005]** According to a first aspect of the present invention, a voice data processing method based on an on-board voice AI is provided. The method includes: acquiring an audio feature and a lyrics feature of a target song; generating a singing voice data of the on-board voice AI for the target song based on the audio feature and the lyrics feature; broadcasting the target song and the singing voice data of the on-board voice AI at the same time; and collecting and broadcasting an audio data of a target user in real time.

**[0006]** According to a second aspect of the present invention, a voice data processing apparatus based on

an on-board voice AI is provided, including: an acquisition unit, configured to acquire an audio feature and a lyrics feature of a target song; a generation unit, configured to generate a singing voice data of an on-board voice AI for the target song based on the audio feature and the lyrics feature; a loudspeaking unit, configured to broadcast the target song and the singing voice data of the on-board voice AI at the same time; and a collection and broadcasting unit, configured to collect and broadcast an audio data of a target user in real time.

**[0007]** According to a third aspect of the present invention, a computer-readable storage medium is provided. The computer-readable storage medium includes a stored program, which when executed by a processor causes the processor to implement the above voice data processing method based on the on-board voice AI.

**[0008]** According to a fourth aspect of the present invention, an electronic device is provided, including at least one processor and at least one memory connected to the processor; the processor is configured to call program instructions in the memory and execute the above voice data processing method based on the on-board voice AI.

**[0009]** The above description is only an overview of technical solutions of the present invention. In order to make technical means of the present invention understood more clearly to be implemented according to the content of the description, and in order to make the above and other objects, features and advantages of the present invention more obvious and understandable, the embodiments of the present invention are specifically listed below.

### BRIEF DESCRIPTION OF DRAWINGS

**[0010]** Various other advantages and benefits will become apparent to those skilled in the art upon reading the following detailed description of the embodiments. The accompanying drawings are only used to illustrate some embodiments of the present invention and are not considered to be limitations of the present invention. Throughout the accompanying drawings, the same reference characters are used to represent the same components. In the accompanying drawings:

FIG. 1 shows a schematic flow chart of a voice data processing method based on an on-board voice AI according to some embodiments of the present invention;

FIG. 2 shows a structural block diagram of a voice data processing apparatus based on an on-board voice AI according to some embodiments of the present invention; and

FIG. 3 shows a structural block diagram of an electronic device according to some embodiments of the present invention.

## DESCRIPTION OF EMBODIMENTS

**[0011]** Exemplary embodiments of the present invention will be described in more detail below with reference to the accompanying drawings. Although exemplary embodiments of the present invention are shown in the accompanying drawings, it should be understood that the present invention may be implemented in various forms and should not be limited to the embodiments set forth herein. Rather, these embodiments are provided to enable a thorough understanding of the present invention, and to fully convey a scope of the present invention to those skilled in the art.

**[0012]** At present, a user cannot use a karaoke software to chorus or duet with an AI assistant, and the users can only shift into an original vocal for chorusing, that is, the user cannot chorus favorite songs with the AI assistant at the same time, and in terms of singing entertainment, an interactivity with the AI assistant is relatively low, resulting in a low competitiveness of intelligent cockpit products equipped with this type of karaoke software. For this reason, embodiments of the present invention provide a voice data processing method based on an on-board voice AI. As shown in FIG. 1, the method may include step S101 to step S104.

**[0013]** In step S101, an audio feature and a lyrics feature of a target song are acquired.

**[0014]** In some embodiments, the above practical application scenario may be that when a vehicle is in a startup state, the user first selects a chorus mode by operating a karaoke application installed in an In-Vehicle Infotainment of the vehicle. There are a plurality of songs in the chorus mode that have been classified and the chorus mode may be a duet mode or a sing-along mode. Then the user selects the song to be chorused.

**[0015]** It can be understood that the target song may be a song for chorusing selected by the user, and the In-Vehicle Infotainment can acquire the audio feature and the lyrics feature of a selected song for chorusing. The audio feature may be a phoneme information, an intonation information, a prosodic boundary text information, a note information, a beat information, a slur and score information and so on of the selected song for chorusing. The lyrics feature may be a lyric information stored in a text data corresponding to the selected song for chorusing, or may be the lyric information obtained by analyzing the audio data corresponding to the selected song for chorusing.

**[0016]** In step S102, a singing voice data of the on-board voice AI for the target song is generated based on the audio feature and the lyrics feature.

**[0017]** It can be understood that the AI assistant is an on-board voice AI application configured in the In-vehicle infotainment. During an implementation procedure, the singing voice data of the on-board voice AI for the target song can be generated based on the audio feature, the lyrics feature and the on-board voice AI application, for example the singing voice data of the on-board voice AI is

generated based on the audio feature, the lyrics feature, and a synthesis solution model of an AI singing voice configured in the AI assistant.

**[0018]** It should be noted that the AI assistant can generate the singing voice data of the on-board voice AI for the target song by the synthesis solution model of the AI singing voice configured with a deep learning neural network algorithm, according to the phoneme information, the intonation information, the prosodic boundary text information, the note information, the beat information, the slur and score information and the lyrics information of the target song acquired in step S101. In some embodiments, this synthesis solution model of the AI singing voice is configured after a model-training by using massive data.

**[0019]** Taking that the choral song selected by the user in the In-Vehicle Infotainment is "Good Luck Comes" as an example, the AI assistant can generate the singing voice data of the on-board voice AI for the song "Good Luck Comes", based on the lyrics information of the song "Good Luck Comes"; the phoneme information, the intonation information, and the prosodic boundary text information which correspond to the text in the lyrics information; the note information, the beat information, the slur and score information of an accompaniment audio of the song "Good Luck Comes"; and the synthesis solution model of the AI singing voice.

**[0020]** It should be understood that a phoneme is a smallest phonetic unit divided according to natural properties of a voice. If an analysis is based on pronunciation actions in a syllable, one pronunciation action constitutes one phoneme. The phonemes are divided into two categories: vowels and consonants. For example, Chinese syllable 啊 (ā) has only one phoneme; Chinese syllable 爱 (ài) has two phonemes; and Chinese syllable 代 (dài) has three phonemes.

**[0021]** An intonation refers to changes of tones of a language, and is high or low of a sound that is inherent in Chinese syllables and has different meanings. A pitch of the intonation is relative, not absolute. The changes in intonation are in a sliding motion rather than in a jumping movement from one scale to another. A five-level tone mark is usually used to present whether the intonation is high or low.

**[0022]** Prosodic boundaries play an important role on two indexes: a naturalness and accuracy of a language expression. In people's communication, pauses among sentences are the prosodic boundaries.

**[0023]** In step S103, the target song and the singing voice data of the on-board voice AI are broadcasted at the same time.

**[0024]** In some embodiments, the selected target song and the singing voice data of the on-board voice AI can be processed in terms of audio based on the karaoke application in the In-vehicle infotainment, and then output through a loudspeaker device in the In-vehicle infotainment. During an implementation procedure, an output of

the target song with the original vocal or the output of the target song with only an accompaniment can be achieved correspondingly by a way of turning on or turning off the original vocal.

**[0025]** By broadcasting the singing voice data of the on-board voice AI and the target song with the original vocal at the same time, or broadcasting the singing voice data of the on-board voice AI and the target song with only the accompaniment at the same time, a good foundation can be laid for the chorus of subsequent users, so that an overall audio output experience can be made better. In addition, by a way of freely turning on the original vocal, three kinds of sound rays can be chorused at the same time when the user subsequently inputs a voice source, thereby enhancing a richness of a singing way.

**[0026]** In step S104, an audio data of the target user is collected and broadcasted in real time.

**[0027]** It should be noted that the In-Vehicle Infotainment can have an external or built-in sound source collection device. While the target song, and the singing voice data of the on-board voice AI are broadcasted, a sound information input by the user currently using the karaoke application through the sound source collection device is collected in real time, and the sound information is used as the audio data of the target user and broadcasted through the loudspeaker device. Of course, the sound information collected can also be processed through an echo cancellation technology, to obtain the audio data of the target user, followed by outputting the audio data through the loudspeaker device.

**[0028]** In some embodiments, when the user uses a karaoke application, an external microphone may be connected to a USB interface of the In-Vehicle Infotainment. After the user selects a song corresponding to his preferred chorus mode, the user can sing at a reasonable angle and within a reasonable collection range of a sound source, to input the sound source from a human voice. At this time, a loudspeaker of the In-Vehicle Infotainment may broadcast the song accompaniment and the singing voice data of the on-board voice AI in step S103. The In-Vehicle Infotainment filters a voice source output by the speakers by using the echo cancellation technology, and conduct a secondary output with low-latency of the sound source from the human voice input by the user and the filtered voice source.

**[0029]** It can be understood that an echo cancellation is a processing method that prevents, by eliminating or removing an audio signal from a far-end sound picked up by a local microphone, the far-end sound from returning back. This kind of removing of the audio signal is done through digital signal processing. A basic principle of the echo cancellation is to establish a voice model of a far-end signal based on a correlation between a loudspeaker signal and a multi-path echo generated by the loudspeaker signal, and use the voice model to estimate the echo, and continuously modify a coefficient of a filter to make an estimated value closer to a real echo. Then, the estimated value of the echo is subtracted from an input signal

of the microphone, thereby achieving a purpose of eliminating the echo.

**[0030]** The above solution can solve the current problem that, inside the vehicle, the user can, based on a mature karaoke software that has been configured in most on-board cockpits, sing by himself or sing with other users through the microphone. This type of application software can process and mix a voice of the user to generate a human voice with a reverberation effect, and then mix the human voice with the song accompaniment to produce a sound of singing. However, the user cannot use this type of application software to chorus or duet with an AI assistant. The user can only shift into an original vocal for chorusing, that is, the user cannot chorus his favorite songs with the AI assistant at the same time. In terms of singing entertainment, an interactivity with the AI assistant is relatively low, resulting in a problem of low competitiveness of the intelligent cockpit products. According to the above method of the present invention, the user can select a song to be chorused by selecting the chorus mode on the karaoke application. The In-Vehicle Infotainment can extract the audio feature and the lyrics feature of the target song selected by the user, generate the singing voice data of the on-board voice AI for the target song, broadcast the target song and the singing voice data of the on-board voice AI, and collect the audio data of the target user in real time, and finally mix and broadcast the above three types of audio, and therefore an effect of improving the interactivity between the user and the AI assistant as well as the entertainment and product competitiveness of the in-vehicle intelligent cockpit can be achieved.

**[0031]** In some embodiments, when the above method is executed, the step S101: acquiring an audio feature and a lyrics feature of a target song, may include a step S201-A or a step S201-B.

**[0032]** In the step S201-A, the audio feature and the lyrics feature of the target song are acquired based on the karaoke application.

**[0033]** It should be noted that the audio feature and the lyrics feature of the target song mentioned in the step S101 can be obtained through direct analysis by the karaoke application. In some embodiments, the karaoke application can directly call an internally cached or downloaded karaoke audio file of the target song and analyze the karaoke audio file, to obtain the audio feature and the lyrics feature of the above audio file.

**[0034]** In the step S201-B, the audio data of the target song is acquired based on the karaoke application; and the audio feature and the lyrics feature of the target song are determined based on the on-board voice AI application and the audio data of the target song.

**[0035]** It should be noted that a difference between the step S201-B and the step S201-A is that in the step S201-B, the internally cached or downloaded karaoke audio file of the target song is transmitted to the on-board voice AI application firstly based on the karaoke application, and then the on-board voice AI application (i.e., the AI assis-

tant) analyzes the above-mentioned karaoke audio file to obtain the audio feature and the lyrics feature of the target song.

**[0036]** The above embodiment designs two ways to analyze the karaoke audio file. During an implementation procedure, it can be decided according to a busy status of a process whether the karaoke audio file of the target song is parsed by the karaoke application or the on-board voice AI application. The resultant In-vehicle infotainment is more smoother during use.

**[0037]** It can be understood that the above methods are divided into two types, A and B. When any one of the two types of methods is executed, a purpose of acquiring a song audio feature and the lyrics feature of the target song can be achieved.

**[0038]** In some embodiments, the above method, when executed, may also include a step S301 that: a sound ray of the AI singing voice of the singing voice data of the on-board voice AI is the same as a sound ray of a voice currently set by the on-board voice AI application.

**[0039]** It should be noted that the sound ray of the on-board voice AI application configured by the In-vehicle infotainment varies, and can be male voices, female voices, dialects and so on. The dialects may be Cantonese, Szechwan dialect, Northeastern dialect and so on. In daily use, the users are often familiar with sounds of the on-board voice AI applications they set, and feel a familiar sense of companionship. The In-Vehicle Infotainment can synchronize, by synchronizing the features of sound rays, the sound ray of the AI singing voice of the singing voice data of the on-board voice AI with the sound ray of the voice currently set by the on-board voice AI application into a same sound ray, such that an experience of chorusing with the voice that accompanies the user every day can be brought to the user; a distance between the on-board voice AI and the user can be shorten; the AI singing voice is no longer a cold companion for the user; and a coordination of an overall chorus is added. Thus a singing experience of the user is improved, and at the same time an overall in-vehicle atmosphere of singing entertainment can be avoided from being disrupted due to a presence of two different sound rays when the In-Vehicle Infotainment outputs certain prompt tones.

**[0040]** In some embodiments, the sound ray set by the on-board voice AI application can be set as a gentle female voice, and the sound ray of the above-mentioned AI singing voice can be set to be synchronized to the sound ray of the voice currently set by the on-board voice AI application. After completion of setting, the singing voice data of the on-board voice AI broadcasted by the loudspeaker according to the sound ray feature of the AI singing voice is based on an audio of the sound ray of the gentle female voice.

**[0041]** In some embodiments, the above method, when executed, may also include a step S401 to step S402.

**[0042]** In the step S401, a singing preference of the target user is determined based on a historical karaoke

data of the target user.

**[0043]** It should be noted that during each singing procedure of the user, the In-vehicle infotainment records a singing habit and emotional ups and downs of the user according to the voice source input by the user through the sound source collection device, and analyzes them by the synthesis solution model of the AI singing voice configured with the deep learning neural network algorithm, to obtain the singing preference of the user, and save this singing preference in the database. The singing preferences can be personalized and classified for storage by setting different storage names.

**[0044]** For example, the users are driver A, passenger B and baby C. The In-vehicle infotainment can remind the user whether to perform a classified storage. After the classified storage is executed, the karaoke application is launched again, and when the user choruses with the AI assistant through the karaoke application, the In-vehicle infotainment can automatically match the sound ray feature of the user with stored singing preferences to enhance the singing experience.

**[0045]** In the step S402, the singing voice data of the on-board voice AI is adjusted according to the singing preference.

**[0046]** It should be noted that the In-Vehicle Infotainment adapts to the singing preference of the user during a chorus procedure according to the singing preference determined in the step S401, thereby making the entire chorus more harmonious and melodious. The above-mentioned singing preference may be an emotional expression of the singing of the user, which may be high-pitched, excited, disappointed, sad and so on, and the emotional expression and a volume for an AI singing voice of the singing voice data of the on-board voice AI are adaptively adjusted according to the above-mentioned singing preferences.

**[0047]** In some embodiments, in the historical karaoke data of a user A recorded in a database, there are mostly sad songs, and singing preferences are mostly low voice volume and moderate intonation. When the user A uses the karaoke application to select the chorus mode to chorus a song "Coral Sea" with the AI assistant, the AI assistant adapts to the singing preferences of the user A and adjusts parameters such as the intonation and the prosodic boundaries and so on to be similar to that of the user A, so as to better complete a collaborative singing of the song.

**[0048]** In some embodiments, when the above method is executed, the step S102: generating a singing voice data of the on-board voice AI for the target song based on the audio feature and the lyrics feature, may include a step S501: transmitting the singing voice data of the on-board voice AI to the Karaoke application based on an Android interface definition language AIDL.

**[0049]** It can be understood that according to the step S102, after the AI assistant, according to the audio feature and the lyrics feature, generates the singing voice data of the on-board voice AI for the target song through

the synthesis solution model of the AI singing voice configured with the deep learning neural network algorithm, the above-mentioned singing voice data of the on-board voice AI needs to be transmitted to the karaoke application and then can be mixed with the target song for broadcasting. The karaoke application and the AI assistant belong to two independent process programs in the In-Vehicle Infotainment, respectively, so the singing voice data of the on-board voice AI needs to be transmitted across processes.

**[0050]** The Android Interface Definition Language (AIDL) is an interface definition language compiled based on Android. Because in Android, different applications respectively run in independent processes, and one application cannot access memory spaces of another application. In order to achieve an inter-process communication, a standard (Peripheral Component Interconnect, PCI) mechanism that defines a local bus is used. Android supports the PCI mechanism, but a serialized data that Android can read is required and AIDL is used to describe the above data.

**[0051]** In some embodiments, the above method, when executed, may also include a step S601: adjusting the sound ray of the voice currently set by the on-board voice AI application based on a sound ray preference of the target user.

**[0052]** It should be noted that different users probably have different preferences for the sound ray of the voice set by the on-board voice AI application. As can be seen from the step S301, the sound ray of the AI singing voice of the singing voice data of the on-board voice AI can be changed with the sound ray of the voice currently set by the on-board voice AI application. Thus a setting of the sound ray of the on-board voice AI application can be adjusted according to the sound ray preference set by the user, to synchronize the sound ray of the AI singing voice to a preferred sound ray of the user. The sound ray may be a gentle female voice, a deep male voice and so on.

**[0053]** In some embodiments, during a procedure that the user A choruses by using the karaoke application, the In-Vehicle Infotainment will identify the user A as a male voice or female voice, and adaptively switch a gender sound ray feature of the AI singing voice. If the sound ray of the AI singing voice has been preset to be the same as the sound ray of the voice currently set by the on-board voice AI application, but the sound ray of the voice currently set by the on-board voice AI application is different from the gender sound ray feature of the AI singing voice adaptively switched, at this time a priority of this step is higher than that of the step S301 by default, and the sound ray of the voice currently set by the on-board voice AI application is overwritten by the gender sound ray feature of the AI singing voice adaptively switched, to achieve a purpose of improving a chorus experience.

**[0054]** In some embodiments, the above method, when executed, may also include a step S701: adaptively adjusting an environmental atmosphere inside the vehi-

cle based on the sound ray of the AI singing voice of the singing voice data of the on-board voice AI and/or the audio feature of the target song.

**[0055]** It should be noted that a lighting and shading device in the vehicle can be adaptively adjusted according to the sound ray of the AI singing voice and/or the audio feature of the target song, to better integrate a scene in the vehicle with an atmosphere of the selected target song, so that the in-vehicle intelligent cockpit is no longer just a carrier of a song for chorusing, but a part of an overall singing atmosphere.

**[0056]** For example, when the user selects the song "Hair is Like Snow" through the karaoke application to chorus with the AI assistant, the In-Vehicle Infotainment controls a sunshade device to automatically close windows of the vehicle, adaptively adjusts a grayscale of glasses of the vehicle, reduces a light transmittance and a saturability of the glasses of the vehicle, adjusts concomitantly an ambient light in the vehicle to ice blue to simulate a snowy atmosphere scene, and adjusts parameters of a low-frequency, a mid-frequency, and a high-frequency of the loudspeaker device, so as to achieve a good singing environment.

**[0057]** It should be noted that, as an implementation of the method shown in the above-mentioned FIG. 1 and several related embodiments, the embodiments of the present invention also provide a voice data processing apparatus based on an on-board voice AI for implementing the method shown by the above-mentioned FIG. 1 and several embodiments. This device embodiment corresponds to the foregoing method embodiment, and can correspondingly implement all contents in the foregoing method embodiment. As shown in FIG. 2, the device may include: an acquisition unit 21, configured to acquire an audio feature and a lyrics feature of a target song; a generation unit 22, configured to generate a singing voice data of an on-board voice AI for the target song based on the audio feature and the lyrics feature; a loudspeaking unit 23, configured to broadcast the target song and the singing voice data of the on-board voice AI at the same time; a collection and broadcasting unit 24, configured to collect and broadcast an audio data of a target user in real time.

**[0058]** In some embodiments, the acquisition unit 21 is also configured to: acquire an audio feature and the lyrics feature of the target song based on the karaoke application; or acquire an audio data of the target song based on the karaoke application; determining the audio feature and the lyrics feature of the target song based on the on-board voice AI application and the audio data of the target song.

**[0059]** In some embodiments, a sound ray of the AI singing voice of the singing voice data of the on-board voice AI is the same as a sound ray of a voice currently set by the on-board voice AI application.

**[0060]** In some embodiments, the device may further include a sound ray adjustment unit (not shown), configured to adjust the sound ray of the voice currently set by

the on-board voice AI application based on a sound ray preference of the target user.

**[0061]** In some embodiments, the generation unit 22 is also configured to: determine a singing preference of the target user based on a historical karaoke data of the target user; and adjust the singing voice data of the on-board voice AI according to the singing preference.

**[0062]** In some embodiments, this device may also include a transmission unit (not shown), configured to transmit the singing voice data of the on-board voice AI to the karaoke application based on an Android interface definition language AIDL.

**[0063]** In some embodiments, this device may also include an atmosphere adjustment unit (not shown), configured to adaptively adjust an environmental atmosphere inside the vehicle based on the sound ray of the AI singing voice of the singing voice data of the on-board voice AI and/or the audio feature of the target song.

**[0064]** By means of the above technical solution, the present invention provides a voice data processing method based on an on-board voice AI, thereby solving a problem of a mature karaoke software currently configured in most on-board cockpits inside the vehicle that: a human voice of a user can be processed and mixed through a microphone to generate the human voice with a reverberation effect and then the human voice is mixed with a song accompaniment to produce a sound of singing, but the user cannot use this type of application software to chorus or duet with an AI assistant; but the user cannot chorus or duet with an AI assistant, and the user can only shift into an original vocal for chorusing, that is, the user cannot chorus his favorite songs with the AI assistant at the same time. Therefore in terms of singing entertainment, an interactivity with the AI assistant is relatively low, resulting in a problem of low competitiveness of the intelligent cockpit products. The present invention, by extracting the audio feature and the lyrics feature of the target song selected by the user, can generate the singing voice data of the on-board voice AI for the target song, broadcast the target song and the singing voice data of the on-board voice AI, and collect the audio data of the target user in real time, and finally mix and broadcast the above three types of audio, and therefore an effect of improving the interactivity between the user and the AI assistant as well as the entertainment and product competitiveness of the in-vehicle intelligent cockpit can be achieved.

**[0065]** A processor contains a core, which retrieves a corresponding program unit from a memory. One or more cores can be set. By adjusting core parameters, a voice data processing method based on an on-board voice AI can be implemented to solve the problem in the existing technology that the user cannot chorus or duet with the AI assistant, and the user can only shift into an original vocal for chorusing, that is, the user cannot chorus his favorite songs with the AI assistant at the same time; in terms of singing entertainment, an interactivity with the AI assistant is relatively low,

**[0066]** Embodiments of the present invention provide a storage medium on which a program is stored, and the program, when executed by the processor, implements the voice data processing method based on the on-board voice AI.

**[0067]** Embodiments of the present invention provide a processor, the processor is configured to run a program, and the program, when running, executes the voice data processing method based on the on-board voice AI.

**[0068]** The embodiment of the present invention provides an electronic device 30. As shown in FIG. 3, the electronic device includes at least one processor 31, and at least one memory 32 and bus 33 connected to the processor; in which a mutual communication of the processor 31 and the memory 32 is achieved through the bus 33; the processor 31 is configured to call program instructions in the memory to execute the above-mentioned voice data processing method based on the on-board voice AI.

**[0069]** The electronic device in the present invention may be a server, PC, PAD, mobile phone and so on.

**[0070]** The present invention also provides a computer program product, which, when executed on a data processing device, is adapted to execute a program initialized with the following steps of method: acquiring an audio feature and a lyrics feature of a target song; and generating a singing voice data of the on-board voice AI for the target song based on the audio feature and the lyrics feature; broadcasting the target song and the singing voice data of the on-board voice AI at the same time; and collecting and broadcasting an audio data of a target user in real time.

**[0071]** In some embodiments, acquiring the audio feature and the lyrics feature of the target song includes: acquiring a song audio feature and the lyrics feature of the target song based on the karaoke application; or acquiring the audio data of the target song based on the karaoke application; determining the audio feature and the lyrics feature of the target song based on the on-board voice AI application and the audio data of the target song.

**[0072]** In some embodiments, the above method further includes: a sound ray of the AI singing voice of the singing voice data of the on-board voice AI is the same as a sound ray of a voice currently set by the on-board voice AI application.

**[0073]** In some embodiments, the above method further includes: determining a singing preference of the target user based on a historical karaoke data of the target user; and adjusting the singing voice data of the on-board voice AI according to the singing preference.

**[0074]** In some embodiments, the above method further includes: the singing voice data of the on-board voice AI is transmitted to the Karaoke application based on an Android interface definition language AIDL.

**[0075]** In some embodiments, the above method further includes: the sound ray of the voice currently

set by the on-board voice AI application is adjusted based on a sound ray preference of the target user.

**[0076]** In some embodiments, the above method further includes: an environmental atmosphere inside the vehicle is adaptively adjusted based on the sound ray feature of the AI singing voice of the singing voice data of the on-board voice AI and/or the song audio feature of the target song.

**[0077]** By means of the above technical solution, the present invention provides a voice data processing method based on an on-board voice AI, thereby solving a problem of a mature karaoke software currently configured in most on-board cockpits inside the vehicle that: a human voice of a user can be processed and mixed through a microphone to generate the human voice with a reverberation effect and then the human voice is mixed with a song accompaniment to produce a sound of singing, but the user cannot use this type of application software to chorus or duet with an AI assistant; but the user cannot chorus or duet with an AI assistant, and the user can only shift into an original vocal for chorusing, that is, the user cannot chorus his favorite songs with the AI assistant at the same time. and therefore in terms of singing entertainment, an interactivity with the AI assistant is relatively low, resulting in a problem of low competitiveness of the intelligent cockpit products. The present invention, by extracting the audio feature and the lyrics feature of the target song selected by the user, can generate the singing voice data of the on-board voice AI for the target song, broadcast the target song and the singing voice data of the on-board voice AI, and collect the audio data of the target user in real time, and finally mix and broadcast the above three types of audio, and therefore an effect of improving the interactivity between the user and the AI assistant as well as the entertainment and product competitiveness of the in-vehicle intelligent cockpit can be achieved.

**[0078]** The present disclosure is described with reference to flowcharts and/or block diagrams of methods, devices (systems), and computer program products according to embodiments of the disclosure. It will be understood that each process and/or block in the flowcharts and/or block diagrams, and combinations of processes and/or blocks in the flowchart and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, embedded computer, or other programmable data processing apparatus to produce a device, such that the device produced by the instructions executed by a processor of the computer or other programmable data processing apparatus can implement the functions specified in a process or processes in the flowcharts and/or in a block or blocks in the block diagrams.

**[0079]** In a typical configuration, a device includes one or more processors (CPUs), memory, and buses. The devices may also include input/output interfaces, net-

work interfaces and so on.

**[0080]** The memory may include a non-permanent memory in computer-readable media, a random access memory (RAM) and/or a non-volatile memory, such as a read-only memory (ROM) or a flash memory (flash RAM). The memory includes at least one storage chip. The memory is an example of a computer-readable medium.

**[0081]** The computer-readable media includes both persistent and non-volatile, removable and non-removable media that can be implemented by any method or technology for storage of information. The information may be computer-readable instructions, data structures, modules of programs, or other data. Examples of a computer storage media include but are not limited to phase change memory (PRAM), static random access memory (SRAM), dynamic random access memory (DRAM), other types of random access memory (RAM), read-only memory (ROM), electrically erasable programmable read-only memory (EEPROM), flash memory or other memory technology, compact disc and read-only memory (CD-ROM), digital versatile disc (DVD) or other optical storage, magnetic tape cassette, tape magnetic disk storage or other magnetic storage devices or any other non-transmission medium which can be used to store information that can be accessed by a computing device. As defined in the present invention, the computer-readable media does not include a computer readable transitory media, such as modulated data signals and carrier waves.

**[0082]** It should also be noted that the terms "comprise", "include" or any other variation thereof are intended to cover a non-exclusive inclusion, such that a procedure, method, article, or device that includes a series of elements not only includes those elements, but also includes other elements not expressly listed or elements are inherent to the procedure, method, article or device. Without further limitation, an element qualified by a statement "comprises a ..." does not exclude a presence of additional identical elements in the procedure, method, article, or device that includes the element.

**[0083]** Those skilled in the art will appreciate that embodiments of the present invention may be provided as methods, systems, or computer program products. Accordingly, the present invention may take the form of an entirely hardware embodiment, an entirely software embodiment, or an embodiment that combines software and hardware aspects. Furthermore, the present invention may take the form of a computer program product embodied on one or more computer-usable storage media (including, but not limited to, disk storage, CD-ROM, optical storage, etc.) having a computer-usable program code embodied therein.

**[0084]** The above are only embodiments of the present invention and are not intended to limit the present invention. Various modifications and variations may occur to the present invention for those skilled in the art. Any modifications, equivalent substitutions, improvements and so on made within a spirit and principle of the present



invention shall be included in the scope sought for by the present invention.

## Claims

1. A voice data processing method based on an on-board voice AI, **characterized by** comprising:

acquiring an audio feature and a lyrics feature of a target song;  
generating a singing voice data of the on-board voice AI for the target song based on the audio feature and the lyrics feature;  
broadcasting the target song and the singing voice data of the on-board voice AI at the same time; and  
collecting and broadcasting an audio data of a target user in real time.

2. The method according to claim 1, wherein the acquiring the audio feature and the lyrics feature of the target song comprises:

acquiring the audio feature and the lyrics feature of the target song based on the karaoke application; or  
acquiring an audio data of the target song based on the karaoke application; and  
determining the audio feature and the lyrics feature of the target song based on an on-board voice AI application and the audio data of the target song.

3. The method according to claim 2, wherein a sound ray of an AI singing voice of the singing voice data of the on-board voice AI is the same as a sound ray of a voice currently set by the on-board voice AI application.

4. The method according to claim 2, further comprising: adjusting the sound ray of the voice currently set by the on-board voice AI application based on a sound ray preference of the target user.

5. The method according to claim 1, further comprising:

determining a singing preference of the target user based on a historical karaoke data of the target user; and  
adjusting the singing voice data of the on-board voice AI according to the singing preference.

6. The method according to claim 1, further comprising: transmitting the singing voice data of the on-board voice AI to a Karaoke application based on an Android interface definition language AIDL.

7. The method according to claim 1, further comprising: adaptively adjusting an environmental atmosphere inside a vehicle based on a sound ray of an AI singing voice of the singing voice data of the on-board voice AI and/or the audio feature of the target song.

8. A voice data processing apparatus based on an on-board voice AI, **characterized by** comprising:

an acquisition unit, configured to acquire an audio feature and a lyrics feature of a target song;  
a generation unit, configured to generate a singing voice data of the on-board voice AI for the target song based on the audio feature and the lyrics feature  
a loudspeaking unit, configured to broadcast the target song and the singing voice data of the on-board voice AI at the same time; and  
a collection and broadcasting unit, configured to collect and broadcast an audio data of a target user in real time.

9. A computer-readable storage medium, **characterized by** comprising a stored program, which, when executed by a processor, causes the processor to implement the voice data processing method based on the on-board voice AI according to any one of claims 1 to 7.

10. An electronic device, **characterized by** comprising at least one processor and at least one memory connected to the processor; the processor is configured to call program instructions in the memory and execute the voice data processing method based on the on-board voice AI according to any one of claims 1 to 7.

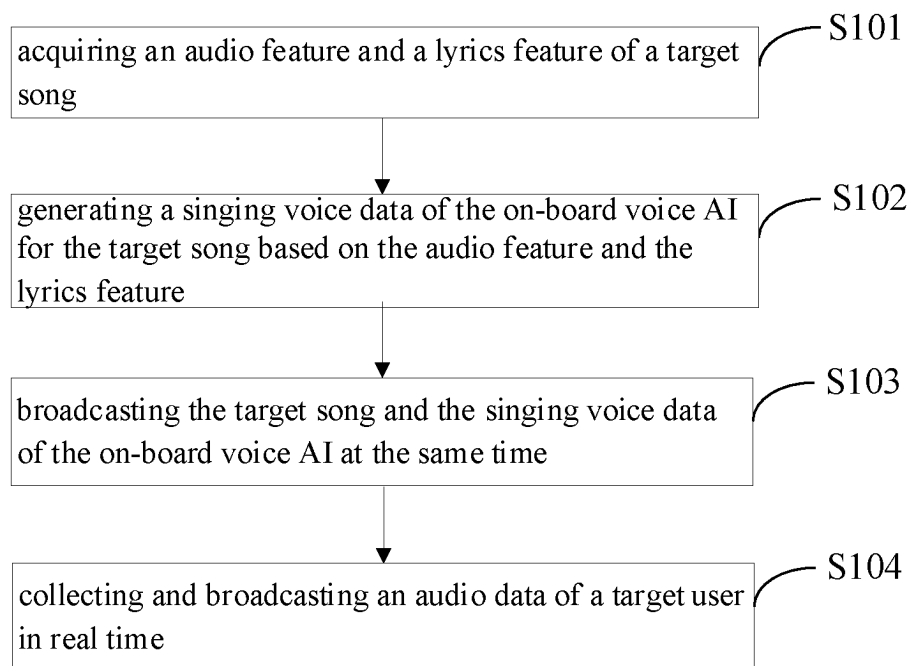


FIG. 1

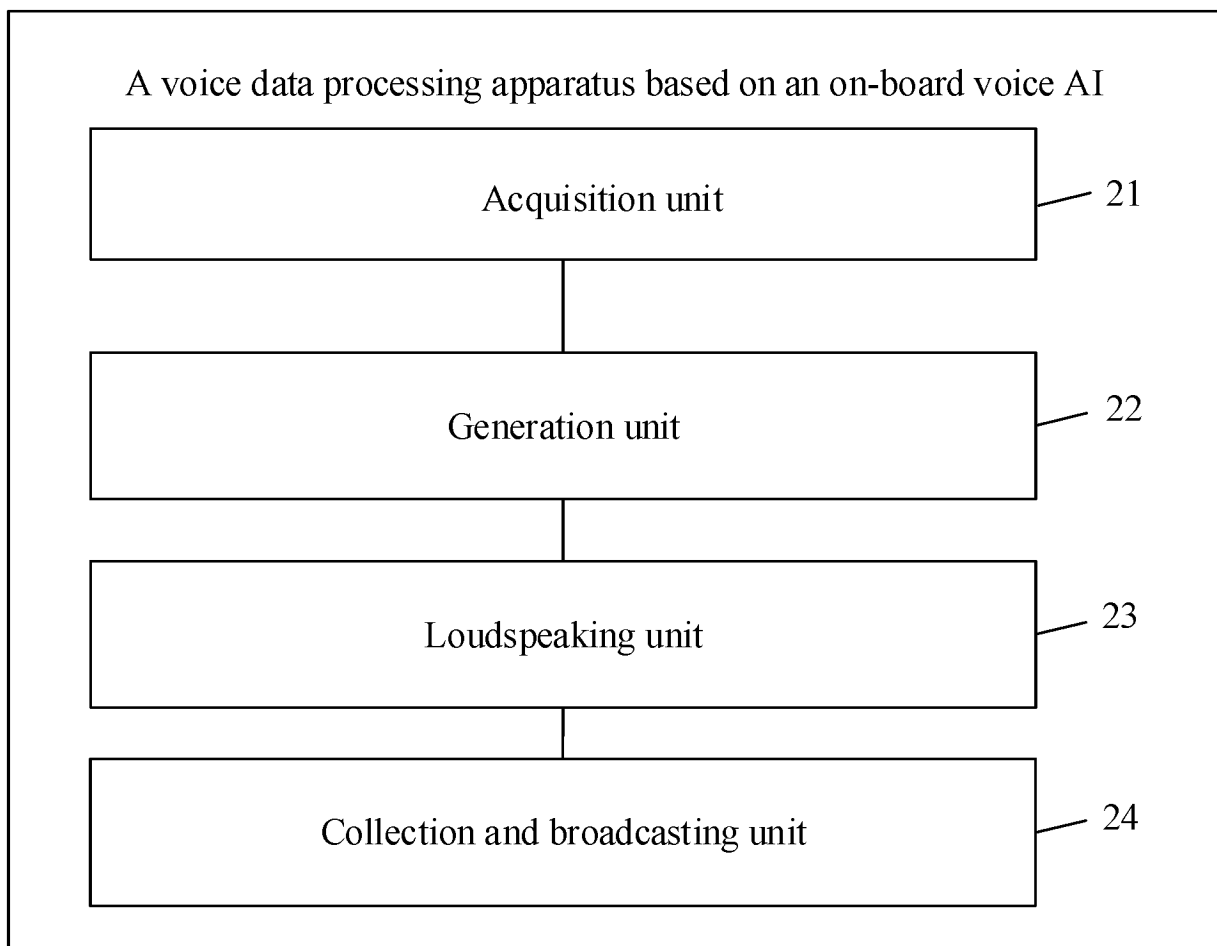


FIG. 2

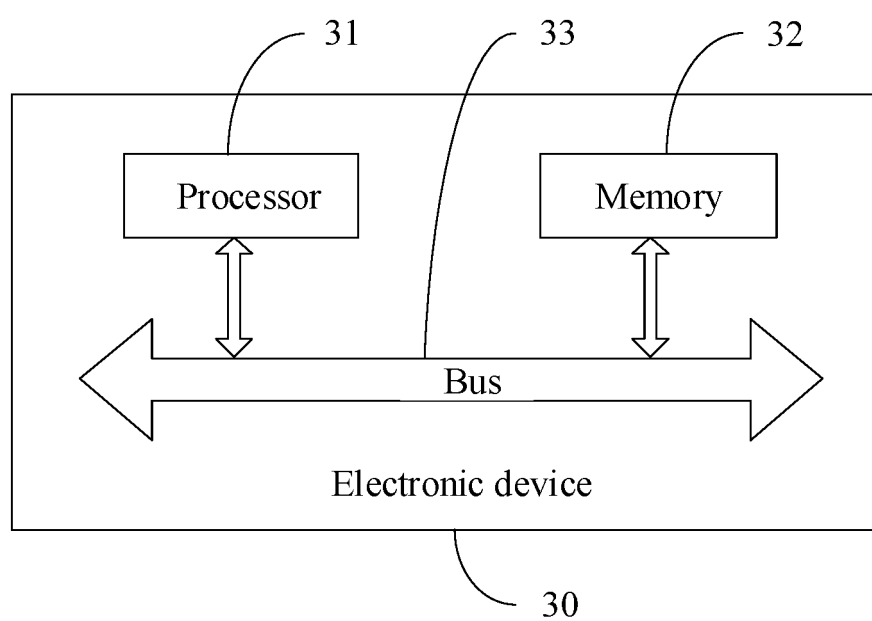


FIG. 3

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/105292

## A. CLASSIFICATION OF SUBJECT MATTER

G10L13/02(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC:G10L13

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS, CNTXT, CNKI: 语音, 音频, 车载, 汽车, 歌曲, 唱歌, 演唱, 对唱, 合唱, K歌, 混合, 人工智能, 机器人, 助手; VEN, USTXT, EPTXT, ELSEVIER: voice, frequency, vehicle, car, song, music, sing, chorus, mix, AI, artificial intelligence, robot, assistant

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 115938340 A (VOYAH AUTOMOBILE TECHNOLOGY CO., LTD.) 07 April 2023 (2023-04-07) description, paragraphs 38-114, and figures 1-3	1-10
A	CN 113225716 A (BEIJING SAIBIN SCIENCE & TECHNOLOGY CO., LTD.) 06 August 2021 (2021-08-06) description, paragraphs 97-227, and figures 1-15	1-10
A	CN 109003623 A (AIWAYS AUTOMOBILE CO., LTD.) 14 December 2018 (2018-12-14) entire document	1-10
A	CN 111754965 A (BYD CO., LTD.) 09 October 2020 (2020-10-09) entire document	1-10
A	CN 107957908 A (SHENZHEN SKYWORTH DIGITAL TECHNOLOGY CO., LTD.) 24 April 2018 (2018-04-24) entire document	1-10
A	US 5235124 A (PIONEER ELECTRONIC CORP.) 10 August 1993 (1993-08-10) entire document	1-10

☒ Further documents are listed in the continuation of Box C.
 ☒ See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&amp;” document member of the same patent family

Date of the actual completion of the international search

20 September 2023

Date of mailing of the international search report

09 October 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/  
CN)  
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,  
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No. <b>PCT/CN2023/105292</b>
-----------------------------------------------------------

5  
  
  
10  
  
15  
  
20  
  
25  
  
30  
  
35  
  
40  
  
45  
  
50  
  
55

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5518408 A (YAMAHA CORP.) 21 May 1996 (1996-05-21) entire document	1-10

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2023/105292**

5

10

15

20

25

30

35

40

45

50

55

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	115938340	A	07 April 2023	None			
CN	113225716	A	06 August 2021	None			
CN	109003623	A	14 December 2018	None			
CN	111754965	A	09 October 2020	None			
CN	107957908	A	24 April 2018	None			
US	5235124	A	10 August 1993	EP	0509812	A2	21 October 1992
				CA	2066018	A1	20 October 1992
				JPH	05341793	A	24 December 1993
US	5518408	A	21 May 1996	JP	H06289872	A	18 October 1994

**REFERENCES CITED IN THE DESCRIPTION**

*This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.*

**Patent documents cited in the description**

- CN 202211335986 [0001]