

[Document Name] Description

[Title of Invention] SONG CREATION ASSISTANCE PROGRAM,
SONG CREATION ASSISTANCE DEVICE, SONG CREATION ASSISTANCE
METHOD, MUSICAL REPRODUCTION DEVICE, AND PERFORMANCE
AUDIO CREATION ASSISTANCE PROGRAM

[Technical Field]

[0001] The present invention relates to a song creation assistance program, a song creation assistance device, a song creation assistance method, a musical reproduction device, and a performance audio creation assistance program.

[Background Art]

[0002] Conventionally, when a new singer debuts as a singer, they need a lot of training of singing ability, such as undergoing so much voice training. However, no matter how much training of singing ability they have undergone, it has been extremely difficult for singers with immature singing ability to debut as a singer in reality unless they have excellent singing ability, such as being evaluated in auditions and other screening as having professional-level ability.

[0003] In addition, producing a song of a singer (e.g., the debut song of a new singer) currently requires a lot of time, effort, and money every time production is

performed. For example, it takes many hours to record a single song, and the following pitch adjustment and editing work for mixing the singing voice with the melody and accompaniment also take several days. Further, there are many expenses for a recording studio, recording equipment, expensive personal computers, editing software, and the like, as well as much effort and high labor cost for those involved (e.g., a singer, a music producer, and technical staff).

[0004] On the other hand, due to the recent rapid development of AI technology, there have been known techniques related to singing voice quality conversion. For example, the invention described in Patent Literature 1 describes a singing voice quality conversion program and a singing voice quality conversion apparatus that, in karaoke or the like, use a vocal model of a target specified by a singing user to convert the voice quality of the user who sings based on a song reserved by the singing user into the voice quality of the target and perform output.

[Citation List]

[Patent Literature]

[Patent Literature 1]

[0005] Japanese Patent Laid-Open No. 2023-013684

[Summary of Invention]

[Technical Problem]

[0006] However, since the invention described in Patent Literature 1 converts the voice quality of a singing user into the voice quality of a specific singer called a target, it does not make the singing ability of the singing user leap to professional-level singing like a specific singer. Therefore, if the singing ability of the singing user is immature, the singing ability remains immature even if it is converted into the voice quality of a specific singer.

[0007] As mentioned above, considering that, for example, when a new singer debuts as a singer, it is extremely difficult for a singer with immature singing ability to debut as a singer in reality no matter how much they wish and recording a song requires a lot of time, effort, and money every time recording is performed, the present inventors have considered that converting the vocal singing voice of a professional singer into the singing voice of a new singer and outputting the singing voice of the new singer by AI-based singing voice quality conversion technique enables anyone to obtain professional-level "singing voice" with their own voice quality, regardless of their singing ability, of course with no need for voice training or other training of singing ability. In addition, the present inventors have considered that since the present invention can greatly reduce the time, effort, and money for song creation

compared to conventional techniques, anyone can debut as a singer in a short period of time with low cost.

[0008] The present invention has been proposed in light of the above points and aims to expand debut opportunities for those who have immature skills such as singing skills or acting skills in one aspect.

[Solution to Problem]

[0009] In order to solve the above problems, a song creation assistance program according to the present invention is for functioning as: first singing voice recording means for recording singing voice data of a first singer; voice quality conversion model creation means for creating a voice quality conversion model of the first singer based on the singing voice data of the first singer; second singing voice recording means for recording first language singing voice data of a song sung by a second singer in a first language, and recording second language singing voice data of the song sung by the second singer or a third singer in a second language; and voice quality conversion means for converting each of the first language singing voice data and the second language singing voice data into voice quality of the first singer by using the voice quality conversion model of the first singer.

[0010] In order to solve the above problems, a musical reproduction device according to the present invention

includes: first singing voice recording means for recording singing voice data of a first singer; voice quality conversion model creation means for creating a voice quality conversion model of the first singer based on singing voice of the first singer; second singing voice recording means for recording first language singing voice data of a song sung by a second singer in a first language, and recording second language singing voice data of the song sung by the second singer or a third singer in a second language; and voice quality conversion means for converting each of the first language singing voice data and the second language singing voice data into voice quality of the first singer by using the voice quality conversion model of the first singer.

[0011] In order to solve the above problems, in a performance audio creation assistance method according to the present invention, a song creation assistance device executes: a first singing voice recording step of recording singing voice data of a first singer; a voice quality conversion model creation step of creating a voice quality conversion model of the first singer based on singing voice of the first singer; a second singing voice recording step of recording first language singing voice data of a song sung by a second singer in a first language, and recording second language singing voice data of the song sung by the second singer or a third

singer in a second language; and a voice quality conversion step of converting each of the first language singing voice data and the second language singing voice data into voice quality of the first singer by using the voice quality conversion model of the first singer.

[Advantageous Effects of Invention]

[0012] According to the embodiments of the present invention, it is possible to expand debut opportunities for those who have immature skills such as singing skills or acting skills in one aspect.

[Brief Description of Drawings]

[0013]

[Figure 1] Figure 1 is a diagram showing an example of the system configuration of a song creation assistance system according to this embodiment.

[Figure 2] Figure 2 is a diagram showing an example of the hardware configuration of a song creation assistance server according to this embodiment.

[Figure 3] Figure 3 is a diagram showing an example of the software configuration of the song creation assistance server according to this embodiment.

[Figure 4] Figure 4 is a diagram showing an example of the data configuration of DBs according to this embodiment.

[Figure 5] Figure 5 is a diagram showing an example of the data configuration of DBs according to this embodiment.

[Figure 6] Figure 6 is a flowchart 1 related to the song creation assistance system according to this embodiment.

[Figure 7] Figure 7 is a flowchart 2 related to the song creation assistance system according to this embodiment.

[Figure 8] Figure 8 is a flowchart 3 related to the song creation assistance system according to this embodiment.

[Figure 9] Figure 9 is a diagram showing an example of the appearance of a musical reproduction device according to this embodiment.

[Figure 10] Figure 10 is a diagram showing an example of the software configuration of a performance audio creation assistance server according to this embodiment.

[Figure 11] Figure 11 is a flowchart 1 related to a performance audio and video creation assistance system according to this embodiment.

[Figure 12] Figure 12 is a flowchart 2 related to the performance audio and video creation assistance system according to this embodiment.

[Figure 13] Figure 13 is a flowchart 3 related to the performance audio and video creation assistance system according to this embodiment.

[Figure 14] Figure 14 is a diagram showing an example of the system configuration of a song creation assistance system according to this embodiment.

[Figure 15] Figure 15 is a diagram showing an example of the software configuration of a song creation assistance server according to this embodiment.

[Description of Embodiments]

[0014] Embodiments of the present invention will be described in detail with reference to the drawings.

[Embodiment 1]

<System Configuration>

Figure 1 is a diagram showing an example of the system configuration of a song creation assistance system according to this embodiment. A song creation assistance system 100 in Figure 1 includes a song creation assistance server 10, a PC 20, a professional singer's microphone 30, a new singer's microphone 40, and a CD writer 50.

[0015] The song creation assistance server 10 is a server apparatus that prepares beforehand the vocal singing voice of a singer with high singing skills (hereinafter referred to as a professional singer), extracts voice quality features of a target singer (hereinafter referred to as a new singer who will debut as a singer) by using a voice quality conversion technique, and converts the vocal singing voice of the professional singer into the voice quality of the target new singer, thereby assisting the creation of a song for the target singer. The song creation assistance server 10 makes edits, such as pitch

adjustment for matching the singing voice after the voice quality conversion to the basic frequency of the singing voice of the target singer, reverbing the vocal (and harmonic vocals) after the voice quality conversion and the pitch adjustment, stereo conversion and noise removal, and mixes it with the melody and accompaniment to complete the final song (e.g., the debut song of a new singer).

[0016] The PC (Personal Computer) 20 is an operating terminal for accessing the song creation assistance server 10 through a network including the Internet and performing operations related to song creation. The PC 20 is implemented, for example, by a personal computer, a tablet terminal or a smartphone, and a predetermined application program (e.g. a general-purpose web browser or a dedicated app) is installed thereon. Note that, by installing a program that implements the functions of the song creation assistance server 10 on the PC 20, the PC 20 and the song creation assistance server 10 may be implemented as an integrated apparatus.

[0017] For the purpose of preparing beforehand high-quality vocal singing voice of a professional singer, the professional singer records, or uses the microphone 30, and stores the singing voice of the professional singer by singing a song itself that is desired to be released as a song of a new singer (e.g., the debut song of the new singer). In addition, as the singing voice of the

professional singer, singing voice recorded in a recording studio in the same way as before may be used.

[0018] For the purpose of creating a voice quality conversion model of a new singer who will debut as a singer, the new singer uses the microphone 40 to record and store singing voice (samples) of the new singer. The songs from which the voice quality conversion model is created are predetermined songs and are desirably about ten songs, but they may be one song as long as the voice quality features of the new singer can be extracted. The predetermined songs may include the same song as the debut song of the new singer, or may not necessarily include it as long as the voice quality features of the new singer can be extracted. In addition, as long as the voice quality features of the new singer can be extracted, it is not necessary to use the singing voice of the new singer, for example, a conversational voice or a reading voice may be used.

[0019] The CD writer 50 is an apparatus that burns the completed final song (e.g., the debut song of the new singer) onto a CD (Compact Disc) as an example of a storage medium. The CD of the song produced in this way will be sold as the debut song of the new singer.

[0020] Since the song creation assistance system 100 according to this embodiment can convert the singing voice of a professional singer into singing voice with the voice quality of a new singer even for a new singer

with low singing skills, such as an unstable pitch or not being able to sing with vibrato, anyone can debut as a singer in a short period of time at a low cost by selling the converted singing voice as a CD or distributing it as a music data file.

[0021] Note that if the song creation assistance server 10 and the PC 20 are located remotely from each other, such as being connected via the Internet, the professional singer's microphone 30, the new singer's microphone 40, and the CD writer 50 may be connected to the PC 20 side.

[0022] (Hardware Configuration)

Figure 2 is a diagram showing an example of the hardware configuration of the song creation assistance server according to this embodiment. The song creation assistance server 10 includes a CPU (Central Processing Unit) 11, a ROM (Read Only Memory) 12, a RAM (Random Access Memory) 13, an HDD (Hard Disk Drive) 14, a communication device 15, and a GPU (Graphics Processing Unit) 16.

[0023] The CPU 11 and the GPU 16 execute various programs and perform arithmetic processing. The ROM 12 stores programs that are required at startup. The RAM 13 is a working area that temporarily stores the processing on the CPU 11 and stores data. The HDD 14 stores various types of data and programs. The communication device 15

performs data communication with other apparatuses through a communication network or a communication cable.

[0024] (Software Configuration)

Figure 3 is a diagram showing an example of the software configuration of the song creation assistance server according to this embodiment. The song creation assistance server 10 includes a singing voice recording unit 101, a voice quality conversion model creation unit 102, an input reception unit 103, a voice quality conversion unit 104, a singing voice editing unit 105, a mix editing unit 106, a completed song output unit 107, and a storage unit 109 (not shown) as main functional units.

[0025] The singing voice recording unit 101 has the function of recording singing voice data of a first singer (e.g., a new singer) and recording singing voice data of a song sung by a second singer (e.g., a Japanese professional singer) and a third singer (e.g., a foreign professional singer).

[0026] The voice quality conversion model creation unit 102 has the function of creating a first singer voice quality conversion model 109b based on the singing voice data of the first singer.

[0027] The input reception unit 103 receives input from a user for information related to voice quality conversion, such as a new singer, a professional singer, a song, and

a language, when executing the voice quality conversion process.

[0028] The voice quality conversion unit 104 has the function of converting the singing voice data of the song sung by the second singer and the third singer into the voice quality of the first singer by using the first singer voice quality conversion model 109b.

[0029] The singing voice editing unit 105 has the function of performing various editing processes on voice quality-converted singing voice data (singing voice data of a song converted into the voice quality of the first singer), such as pitch adjustment for matching it to the basic frequency of the singing voice of the target new singer, reverbing the vocal (and harmonic vocal) singing voice after the voice quality conversion and the pitch adjustment, stereo conversion and noise removal.

[0030] The mix editing unit 106 has the function of mixing melody or accompaniment data with voice quality-converted singing voice data on which various editing processes are performed.

[0031] The completed song output unit 107 has the function of outputting the voice quality-converted singing voice data (singing voice data of the song) on which various editing processes and mixing are performed as the song data sung by the first singer. As the output, the song data sung by the first singer is stored on an external storage medium (e.g., a CD), or

distributed to an external server via a network including the Internet.

[0032] The storage unit 109 has the function of storing a pre-trained base model 109a, the new singer voice quality conversion model 109b, a new singer information DB 109c, a professional singer singing voice storage DB 109d, a voice quality-converted singing voice storage DB 109e, a song DB 109f, and a completed song DB 109g.

[0033] It goes without saying that these functional units are functional units related to the present invention and the song creation assistance server 10 includes various other functional units necessary for song creation assistance. In addition, each functional unit is implemented by a program that is executed on hardware resources such as a CPU, a GPU, a ROM, a RAM and the like of a computer that constitutes the song creation assistance server 10. These functional units may be read as "means", "modules", "units", or "circuits". In addition, each DB of the storage unit 109 in the song creation assistance server 10 can also be placed on an external storage device on a network. In addition, the above computer programs and application programs may be stored in a computer-readable storage medium.

[0034] (Various DBs)

Figure 4(a) is a diagram showing an example of the data configuration of a new singer information DB according to this embodiment. The new singer information

DB 109c is a DB that pre-registers singer information of new singers, for example, as a learning target whose voice quality is desired to be learned by the voice quality conversion model. The new singer information DB 109c according to this embodiment has data items such as "new singer ID", "name", "gender", and "affiliation" of a new singer. Note that new singers registered in the new singer information DB 109c are assumed to be singers whose voice quality conversion model 109b has already been created.

[0035] Figure 4(b) is a diagram showing an example of the data configuration of a professional singer singing voice storage DB according to this embodiment. The professional singer singing voice storage DB 109d is a DB that pre-registers singer information of a professional singer and singing voice data of a song (e.g., the debut song of a new singer) sung by the professional singer. The professional singer singing voice storage DB 109d according to this embodiment has data items such as "professional singer singing voice ID", "name" of a professional singer, "song ID", "professional singer singing voice data", and "language".

[0036] Among them, "song ID" is an ID for identifying a song sung by the professional singer. It corresponds to a song ID in the song DB 109f. "Professional singer singing voice data" is a singing voice data file recorded by the professional singer previously singing the song

(e.g., the debut song of a new singer). "Language" is the lyric language of the song. Even in the case of the same song, the lyrics may be translated into other languages such as foreign languages. In such a case, the professional singer not only sings the song with Japanese lyrics but also sings the song with lyrics in other languages, resulting in the professional singer singing voice data for the lyrics in other languages.

[0037] In addition, if it is difficult for the professional singer to sing the song in other languages, professional singer singing voice data for the lyrics in other languages may be obtained by other professional singers singing the song in other languages. For example, for the same song ID "M0001", the professional singer singing voice storage DB 109d shown in Figure 4(b) stores professional singer singing voice data (professional singer singing voice ID "P0001") obtained by singing the song with Japanese lyrics with the singing ability of a professional singer M and professional singer singing voice data (professional singer singing voice ID "P0002") obtained by singing the song with English lyrics with the singing ability of a professional singer N who is a native English speaker.

[0038] Figure 5(a) is a diagram showing an example of the data configuration of the voice quality-converted singing voice storage DB according to this embodiment. The voice quality-converted singing voice storage DB 109e is a DB

that stores the professional singer singing voice data of a professional singer that is converted into the voice quality of a target new singer. The voice quality-converted singing voice storage DB 109e according to this embodiment has data items such as "voice quality-converted singing voice ID", "new singer ID", "professional singer singing voice ID", "song ID", and "voice quality-converted singing voice data".

[0039] Among them, "new singer ID" indicates the new singer ID of a target new singer who has the singing voice (voice quality) after the voice quality conversion. "Professional singer singing voice ID" indicates the professional singer singing voice ID of a professional singer who has the singing voice before the voice quality conversion. "Song ID" indicates the song ID of a song sung by the professional singer (e.g., the debut song of a new singer). "Voice quality-converted singing voice data" is singing voice data after the professional singer singing voice data of the professional singer is converted into the voice quality of the target new singer.

[0040] Figure 5(b) is a diagram showing an example of the data configuration of the song DB 109f according to this embodiment. The song DB is a DB that pre-registers information about a song that is desired to be released as a song of a new singer (e.g., the debut song of the new singer). The song DB 109f according to this

embodiment has data items such as "song ID", "song title", and "melody and accompaniment data".

[0041] Among these, "melody and accompaniment data" is voice data such as melody and accompaniment of the song. The singing voice quality-converted singing voice data after being converted into the voice quality of the target new singer is mixed with the melody and accompaniment data to complete the final song.

[0042] Figure 5(c) is a diagram showing an example of the data configuration of the completed song DB according to this embodiment. The completed song DB 109g is a DB that stores information on a song released as a song of a new singer (e.g., the completed debut song of a new singer). The completed song DB 109g according to this embodiment has data items such as "completed song ID", "completed song data", "song title", "new singer ID", and "professional singer singing voice ID".

[0043] Among them, "completed song data" is the song data of the final song (e.g., the completed debut song of a new singer) completed by mixing the singing voice quality-converted singing voice data after being converted into the voice quality of the target new singer with the melody and accompaniment data. "Song title" indicates the title of the song to be released. "New singer ID" indicates the new singer ID of the new singer who will be the singer of the completed song. "Professional singer singing voice ID" indicates the

professional singer singing voice ID of the singing voice of the professional singer who actually sang the song before the voice quality conversion.

[0044] Note that the data configuration of the above DBs according to this embodiment is only an example, and other data items may be included.

[0045] <Information Processing>

(Voice Quality Conversion Model Creation Process)

Figure 6 is a flowchart 1 related to the song creation assistance system according to this embodiment. Causing the CPU or GPU of the song creation assistance server 10 to load and execute a program capable of implementing this flowchart processing can implement each step (hereinafter denoted as "S").

[0046] S1: The song creation assistance server 10 records (obtains) the singing voice of a new singer from which a voice quality conversion model is created through the microphone 40. Specifically, the new singer sings, for example, about ten predetermined songs. Note that for the purpose of description, the new singer is assumed to have poor singing ability.

[0047] S2: The song creation assistance server 10 creates the new singer voice quality conversion model 109b based on features of the recorded singing voice of the new singer. Specifically, it is preferable to use HuBERT (Hidden Unit Bidirectional Encoder Representationd form Transformers) as a basis for creating a voice quality

conversion model. When a voice quality conversion model is created, a pre-trained base model 109a of a general human voice is first created, instead of creating a specific person's voice from scratch. This model learns speech data of many people in many language regions, and has the features of a common human voice. Creating the voice quality conversion model of each individual in such a manner as to superimpose voice information of a person to be learned (e.g., a new singer) on this pre-trained base model 109a makes it possible to generate a voice quality conversion model of a new singer by learning a small amount of voice without the need to prepare a large amount of data of the person to be trained.

[0048] In this way, during training, it is possible to quickly create a highly accurate voice quality conversion model 109b by inputting voice quality feature values (voice feature values) of the new singer who is the conversion destination to the pre-trained base model 109a. During voice quality conversion (inference), voice quality feature values (voice feature values) of the professional singer who is the conversion source are converted by the created voice quality conversion model 109b into the voice quality of the singing voice of the new singer.

[0049] Note that the pre-trained base model 109a according to this embodiment is obtained by previously learning singing voices and voices of many people not

only in Japanese but also in multiple languages, and a base model obtained by learning vowels and consonants of various languages can be used. In addition, a base model specialized for each music genre, such as J-POP, Western music, anime songs, enka, and game songs, can be prepared as the pre-trained base model 109a. For example, in the case of a base model specialized for the anime song genre, incremental learning or transfer learning of singing voice data of professional singers in the anime song genre may further be performed.

[0050] (Professional Singer Singing Voice Storage Process)

As mentioned above, by a professional singer previously singing a song itself that is desired to be released as a song of a new singer (e.g., the debut song of a new singer), the singing voice data of the song sung by the professional singer is recorded and stored in the professional singer singing voice storage DB 109d.

[0051] In addition, a song according to this embodiment has an English lyric version. In such a case, the professional singer not only sings the song with Japanese lyrics but also sings the song with lyrics in other languages, resulting in the professional singer singing voice data for the lyrics in other languages. If it is difficult for the professional singer to sing the song in other languages, professional singer singing voice data for the lyrics in other languages may be obtained by

other professional singers singing the song in other languages. For example, as shown in Figure 4(b), the professional singer singing voice storage DB 109d in this embodiment stores professional singer singing voice data of the same song (song ID: M0001) sung by different professional singers M and N with different language lyrics (in Japanese and English). For example, the professional singer M is a native Japanese singer, and the professional singer N is a native English singer. [0052] In addition, if the song is a song that includes a harmonic vocal, not only the singing voice data of the main vocal sung by a professional singer but also the singing voice data of the harmonic vocal sung by the professional singer or another professional singer is stored separately in the professional singer singing voice storage DB 109d as singing voice data of the professional singer.

[0053] (Voice Quality Conversion Process)

Figure 7 is a flowchart 2 related to the song creation assistance system according to this embodiment.

[0054] S21: The song creation assistance server 10 receives input of a new singer, a professional singer, a song, and a language through the PC 20. For these inputs, it is also possible to provide selective input from among new singers registered in the new singer information DB 109c and professional singers' names,

songs, and languages stored in the professional singer singing voice storage DB 109d.

[0055] S22: When receiving the above inputs in S21, the song creation assistance server 10 converts the corresponding professional singer singing voice data stored in the professional singer singing voice storage DB 109d into the voice quality of the corresponding new singer by using the new singer voice quality conversion model 109b. More specifically, the voice quality feature values (voice feature values) of the professional singer extracted based on the professional singer singing voice data are input to the new singer voice quality conversion model 109b, and the professional singer singing voice data converted into the voice quality of the new singer is obtained as output.

[0056] S23: The song creation assistance server 10 stores the professional singer singing voice data converted into the voice quality of the new singer obtained in S22 in the voice quality-converted singing voice storage DB 109e as voice quality-converted singing voice data together with, for example, the new singer's new singer ID, the professional singer's professional singer ID, the song ID of the song, and the like as shown in Figure 5(a). The professional singer singing voice data after voice quality conversion stored here is vocal voice data that sounds as if the new singer is singing the song with the singing ability equivalent to that of the professional

singer, because its voice quality is that of the new singer.

[0057] In addition, as mentioned above, if the song is a song that includes a harmonic vocal, the singing voice data of the main vocal sung by the professional singer and the singing voice data of the harmonic vocal are stored in the professional singer singing voice storage DB 109d as singing voice data of the professional singer. In this case, when receiving the above inputs in S21, the song creation assistance server 10 can convert the singing voice data of the main vocal stored in the professional singer singing voice storage DB 109d into the voice quality of the new singer and also converts the singing voice data of the harmonic vocal into the voice quality of the new singer by using the new singer voice quality conversion model 109b. In addition, the song creation assistance server 10 stores the professional singer singing voice data of the main vocal and the professional singer singing voice data of the harmonic vocal after being converted into the voice quality of the new singer in the voice quality-converted singing voice storage DB 109e as voice quality-converted singing voice data together with the new singer's new singer ID, the professional singer's professional singer ID, the song ID of the song, and the like.

[0058] (Editing Process)

Figure 8 is a flowchart 3 related to the song creation assistance system according to this embodiment.

S31: The song creation assistance server 10 performs various editing processes on the voice quality-converted singing voice data stored in the voice quality-converted singing voice storage DB 109e. Various editing processes include, for example, pitch adjustment for matching to the basic frequency of the singing voice of the target new singer, reverbing the vocal (and harmonic vocal) singing voice after the voice quality conversion and the pitch adjustment, stereo conversion and noise removal.

[0059] S32: The song creation assistance server 10 mixes the voice quality-converted singing voice data on which various editing processes are performed with, for example, the melody and accompaniment data stored in the song DB 109f in association with the song as shown in Figure 5(b), thereby completing the final song (e.g., the debut song of the new singer).

[0060] S33: The song creation assistance server 10 stores the completed final song (e.g., the debut song of the new singer) in the completed song storage DB 109g as the completed song data, for example, as shown in Figure 5(c). After that, the completed song data is burned on a CD by the CD writer 50 and packaged, or distributed as a music data file, as, for example, the debut song of the new singer.

[0061] The following is an example of attribute information related to the completed song data.

Example 1:

- Completed song ID: M0001
- New singer name: A
- New singer ID: S0001
- Song title: "Rubii Mentali"
- Song ID: M0001
- Professional singer name: M
- Professional singer singing voice ID: P0001
- Language: Japanese

The song title of the debut song of the new singer A is "Rubii Mentali". In fact, it is produced by converting the singing voice obtained by singing the song in Japanese with the singing ability of the professional singer M into the voice quality of the new singer and using the result as the vocal voice.

[0062] As the example 1 shows, the completed song data is produced by converting the singing voice obtained by singing the song with the singing ability of the professional singer into the voice quality of the new singer and using the result as the vocal voice. For this reason, even if a new singer has low singing skills (a new singer is not good at singing), such as an unstable pitch or not being able to sing with vibrato, anyone can debut as a singer in a short period of time at a low cost with a professional singer-level singing voice.

[0063] Example 2:

- Completed song ID: M0002
- New singer name: A
- New singer ID: S0001
- Song title: "Ruby Mentai"
- Song ID: M0001
- Professional singer name: N
- Professional singer singing voice ID: P0002
- Language: English

The song title of the international debut song of the new singer A is "Ruby Mentai" (English lyrics). In fact, it is produced by converting the singing voice obtained by singing the song in English with the singing ability of the professional singer N who is a native English speaker into the voice quality of the new singer and using the result as the vocal voice.

[0064] As the example 2 shows, the singing voice of a professional singer can be converted into the voice quality of a new singer even if it is singing voice in any language as long as the professional singer singing voice data is data of singing voice sung in that language, so even if the new singer cannot speak that language at all or even sing the lyrics in that language, the singing voice of the lyrics in that language sung by the singing voice of the professional singer can be produced with the voice quality of the new singer. As a result, anyone can easily debut as a singer in countries

around the world by converting the singing voice recorded from a professional singer in the foreign language into the voice quality of the new singer and selling CDs or distributing it in the foreign country.

[0065] In addition, even a famous singer with high singing ability who is already well-known in Japan may not be good at a foreign language other than Japanese. However, similarly, the singing voice of a famous singer can be converted into singing voice with lyrics in the foreign language using the singing voice recorded from a professional singer in the foreign language, so famous Japanese singers can easily debut as a singer in countries around the world. For example, if this is applied to a world-famous Japanese anime song sung by a famous professional anime song singer, people around the world can listen to the anime song with lyrics of their own country with the singing voice of the same famous singer.

[0066] <Application Examples>

Figure 9 is a diagram showing an example of the appearance of a musical reproduction device according to this embodiment. The musical reproduction device 60 according to this embodiment includes a CPU (not shown), a RAM (not shown), a storage unit (not shown), a display unit 61, operation units 62-64, a music output unit (not shown), a USB interface 65, and the like.

[0067] The CPU reads out various programs stored in the storage unit based on control signals input from the operation units, deploys them to a work area on the RAM, and executes the programs. The storage unit stores various programs (e.g., a control program and a musical reproduction program) and "completed song data" created by the song creation assistance system 100 according to this embodiment. The display unit 61 displays various screens such as the operating status of the apparatus, the song being played and settings according to a display signal output from the CPU.

[0068] The operation units include a power ON/OFF button 62, play buttons 63 for switching play/stop of "completed song data", and volume control buttons 64 for adjusting the playback volume.

[0069] The music output unit includes a DA converter, an amplifier, a speaker, and the like, and outputs music that is operated to be played from the speaker. The USB interface 65 performs communication according to the USB standard. "Completed song data" is stored in the storage unit, for example, via the USB interface.

[0070] Here, as shown in Figure 9, the play buttons 63 according to this embodiment include pairs of two buttons, and specifically include a total of eight buttons: play buttons 63-1a and 63-1b, play buttons 63-2a and 63-2b, play buttons 63-3a and 63-3b, and play buttons 63-4a and 63-4.

[0071] Then, "completed song data" is associated with each play button one by one, and when the user presses a play button, the musical reproduction device 60 plays "completed song data" associated with the pressed play button. An example of the attribute information related to the "completed song data" associated with each of the play button 63-1a and the play button 63-1b is shown below.

Play button 63-1a:

- Completed song ID: M0001
- New singer name: A
- New singer ID: S0001
- Song title: "Rubii Mentali"
- Song ID: M0001
- Professional singer name: M
- Professional singer singing voice ID: P0001
- Language: Japanese

Play button 63-1b:

- Completed song ID: M0002
- New singer name: A
- New singer ID: S0001
- Song title: "Ruby Mentali"
- Song ID: M0001
- Professional singer name: N
- Professional singer singing voice ID: P0002
- Language: English

[0072] Both of "completed song data" associated with the play button 63-1a and "completed song data" associated with the play button 63-1b are the song title "Rubii Mental" of the debut song of the new singer A. However, "completed song data" associated with the play button 63-1a is produced by converting the singing voice obtained by singing the song (Japanese lyrics) in Japanese with the singing ability of the professional singer M into the voice quality of the new singer A and using the result as the vocal voice. On the other hand, "completed song data" associated with the play button 63-1b is produced by converting the singing voice obtained by singing the song (English lyrics) in English with the singing ability of the professional singer N who is a native English speaker into the voice quality of the new singer A and using the result as the vocal voice.

[0073] In other words, when the user presses the play button 63-1a, the song title "Rubii Mental" (Japanese lyrics) sung by the singing voice of the new singer A is played, while when the user presses the play button 63-1b, the song title "Ruby Mental" (English lyrics) sung by the singing voice of the new singer A is played.

[0074] According to the musical reproduction device 60 in this embodiment, operating the play buttons including pairs of a plurality of (e.g., two) buttons makes it possible to easily switch between the lyric languages to select, play, and listen to one, although they are lyric

languages of the same song sung by the same singer's singing voice. For example, it is possible to listen to the same anime song with the same singer's singing voice with Japanese lyrics or with lyrics in other languages such as English. In addition, it can also be utilized for learning pronunciation in various languages by listening to and comparing the lyrics in various languages while repeatedly switching languages, and singing in imitation while understanding the meaning of the lyrics.

[0075] Note that since the musical reproduction device 60 according to this embodiment has a total of 8 buttons, the user can freely set the association with "completed song data". The following is an example of association between each button and "completed song data" in the case of eight buttons.

- Play button 63-1a: Song A (Japanese lyrics)
- Play button 63-1b: Song A (English lyrics)
- Play button 63-2a: Song B (Japanese lyrics)
- Play button 63-2b: Song B (Chinese lyrics)
- Play button 63-3a: Song C (Japanese lyrics)
- Play button 63-3b: Song C (English lyrics)
- Play button 63-4a: Song C (Chinese lyrics)
- Play button 63-4b: Song C (Thai lyrics)

[0076][Embodiment 2]

The above embodiment 1 has described a mechanism that enables anyone to debut as a singer in a short

period of time at a low cost with a professional singer-level singing voice even if they are a new singer with low singing skills by using the song creation assistance system 100. An embodiment 2 will describe a mechanism that enables anyone to debut as a voice actor in a short period of time at a low cost with a professional voice actor-level performance audio even if they are a new voice actor with low voice acting skills by using a similar mechanism. A voice actor refers to a person who appears and acts in a video work such as anime only with their voice without showing themselves. In a broader sense, narrators are also included.

[0077] Note that, for example, the song creation assistance system 100, the song creation assistance server 10, a new singer, a professional singer, and the like according to the embodiment 1 can be read as a performance audio and video creation assistance system 200, a performance audio creation assistance server 70, a new voice actor, and a professional voice actor, respectively, in the embodiment 2.

[0078] (Software Configuration)

Figure 10 is a diagram showing an example of the software configuration of a performance audio creation assistance server according to this embodiment. The performance audio creation assistance server 70 includes a voice recording unit 701, a voice quality conversion model creation unit 702, an input reception unit 703, a

voice quality conversion unit 704, a performance audio editing unit 705, a mix editing unit 706, a completed video work output unit 707, and a storage unit 709 (not shown) as main functional units.

[0079] The voice recording unit 701 has the function of recording voice data of a first voice actor (e.g., a new voice actor) and recording performance audio data of a song sung by a second voice actor (e.g., a Japanese professional voice actor) and a third voice actor (e.g., a foreign professional voice actor).

[0080] The voice quality conversion model creation unit 702 has the function of creating a first voice actor voice quality conversion model 709b based on the voice data of the first voice actor.

[0081] The input reception unit 703 receives input from a user for information related to voice quality conversion, such as a new voice actor, a professional voice actor, a video work, and a language, when executing the voice quality conversion process.

[0082] The voice quality conversion unit 704 has the function of converting performance audio data of a script acted out by the second voice actor and the third voice actor into the voice quality of the first voice actor by using the first voice actor voice quality conversion model 709b.

[0083] The performance audio editing unit 705 has the function of performing various editing processes on the

voice quality-converted performance audio data (the performance audio data of the script converted into the voice quality of the first voice actor).

[0084] The mix editing unit 706 has the function of mixing video data with the voice quality-converted performance audio data on which various editing processes are performed.

[0085] The completed video work output unit 707 has the function of outputting the voice quality-converted performance audio data (the performance audio data of the script) on which various editing processes and mixing are performed as completed video work data acted out by the first voice actor.

[0086] The storage unit 709 has the function of storing a pre-trained base model 709a, the new voice actor voice quality conversion model 709b, a new voice actor information DB 709c, a professional voice actor performance audio storage DB 709d, a voice quality-converted singing voice storage DB 709e, a video work DB 709f, and a completed video work DB 709g.

[0087] (Voice Quality Conversion Model Creation Process)

Figure 11 is a flowchart 1 related to the performance audio and video creation assistance system according to this embodiment. Causing the CPU or GPU of the performance audio creation assistance server 70 to load and execute a program capable of implementing this flowchart processing can implement each step.

[0088] S51: The performance audio creation assistance server 70 records (obtains) the voice of a new voice actor from which a voice quality conversion model is created through the microphone 40. Specifically, the new voice actor reads, for example, several predetermined scripts. Note that for the purpose of description, the new voice actor is assumed to have poor acting ability.

[0089] S52: The performance audio creation assistance server 70 creates the new voice actor voice quality conversion model 709b based on features of the recorded voice of the new voice actor. Specifically, in order to create a voice quality conversion model, for example, learning and voice quality conversion using well-known voice quality conversion can be applied.

[0090] The existence of a base model that is a pre-training model eliminates the need for preparing a large amount of data on a person to be learned, and a few minutes of voice data makes it possible to accurately learn the voice quality feature values of the target new voice actor. Note that the pre-trained base model 709a according to this embodiment is obtained by previously learning voices of many people not only in Japanese but also in multiple languages, and a base model obtained by learning vowels and consonants of various languages is used.

[0091] (Professional Voice Actor Performance Audio Storage Process)

By a professional voice actor previously acting out lines from the script of a video work desired to be produced as the performance audio of a new voice actor, the performance audio data of the professional voice actor is stored in the professional voice actor performance audio storage DB 709d.

[0092] Note that the professional voice actor not only acts out the script with Japanese lines but also acts out the script with lines in other languages, resulting in the professional voice actor performance audio data for the lines in other languages. Note that if it is difficult for the professional voice actor to act out the script of the video work in other languages, professional voice actor performance audio data for the lines in other languages may be obtained by other professional voice actors acting out the script in other languages.

[0093] (Voice Quality Conversion Process)

Figure 12 is a flowchart 2 related to the performance audio and video creation assistance system according to this embodiment.

S61: The performance audio creation assistance server 70 receives input of a new voice actor, a professional voice actor, a script, and a language through the PC 20. Note that for these inputs, it is also possible to provide selective input from among new voice actors registered in the new voice actor information DB 709c and professional voice actors' names,

scripts, and languages stored in the professional voice actor performance audio storage DB 709d.

[0094] S62: When receiving the above inputs in S61, the performance audio creation assistance server 70 converts the professional voice actor performance audio data stored in the professional voice actor performance audio storage DB 709d into the voice quality of the new voice actor by using the new voice actor voice quality conversion model 709b.

[0095] S63: The performance audio creation assistance server 70 stores the professional voice actor performance audio data converted into the voice quality of the new voice actor obtained in S62 in the voice quality-converted performance audio storage DB 709e as voice quality-converted performance audio data together with the new voice actor's new voice actor ID, the professional voice actor's professional voice actor ID, the script ID of the script, and the like. The performance audio after voice quality conversion stored here is performance audio that sounds as if the new voice actor is acting out the script with the acting ability equivalent to that of the professional voice actor, because its voice quality is that of the new voice actor.

[0096] (Editing Process)

Figure 13 is a flowchart 3 related to the performance audio and video creation assistance system according to this embodiment.

S71: The performance audio creation assistance server 70 performs various editing processes on the voice quality-converted performance audio data stored in the voice quality-converted performance audio storage DB 709e. Various editing processes include, for example, noise removal.

[0097] S72: The performance audio creation assistance server 70 mixes the voice quality-converted performance audio data on which various editing processes are performed with video data, sound effects and BGM stored in the video work DB 709f in association with the script, thereby completing the final video work.

[0098] According to this embodiment, completed video work data is produced by converting the performance audio obtained by reading the script with the acting ability of a professional voice actor into the voice quality of the new voice actor and using the result as the performance audio. For this reason, even if a new voice actor has low acting skills, such as unstable intonation or poor articulation, anyone can debut as a voice actor in a short period of time at a low cost with a professional voice actor-level performance audio.

[0099] In addition, the performance audio of a professional voice actor can be converted into the voice quality of a new voice actor even if it is performance audio of lines in any language as long as the professional voice actor performance audio data is data

of performance audio obtained by reading lines in that language, so even if the new voice actor cannot read the script with lines in that language at all, the performance audio in that language at the level of a professional voice actor can be produced with the voice quality of the new voice actor. As a result, anyone can easily debut as a voice actor in countries around the world by converting the performance audio in a foreign language recorded from a professional voice actor in the foreign language into the voice quality of the new voice actor and selling or distributing video works in the foreign country.

[0100] In addition, even a famous voice actor with high singing ability who is already well-known in Japan may not be good at a foreign language other than Japanese. However, similarly, the voice of a famous voice actor can be converted into performance audio spoken in the foreign language by using the performance audio recorded from a professional voice actor who is fluent in the foreign language, so famous Japanese voice actors can easily debut as a voice actor in countries around the world.

[0101][Embodiment 3]

In an embodiment 3, the song creation assistance system 100 according to the above embodiment 1 is modified to create a voice quality conversion model of a non-existent singer (hereinafter referred to as a virtual

singer) and debut a virtual non-existent person as a singer.

[0102] Figure 14 shows an example of the system configuration of a song creation assistance system according to this embodiment. A song creation assistance system 300 in Figure 14 includes a song creation assistance server 80, the PC 20, the professional singer's microphone 30, a microphone 40 for a plurality of singers, and the CD writer 50.

[0103] The song creation assistance server 10 is a server apparatus that prepares beforehand the vocal singing voice of a singer with high singing skills (hereinafter referred to as a professional singer), extracts voice quality features of a target singer (a non-existent virtual singer) by using a voice quality conversion technique, and converts the vocal singing voice of the professional singer into the voice quality of the target virtual singer, thereby assisting the creation of a song for the target singer. In addition, the song creation assistance server 80 make edits, such as pitch adjustment for matching the singing voice after the voice quality conversion to the basic frequency of the singing voice of the target singer, reverbing the vocal (and harmonic vocals) after the voice quality conversion and the pitch adjustment, stereo conversion and noise removal, and mixes it with the melody and accompaniment to complete

the final song (e.g., the debut song of a virtual singer).

[0104] For the purpose of preparing beforehand vocal singing voice of a professional singer, the professional singer records, or uses the microphone 30, and stores the singing voice of the professional singer by singing a song itself that is desired to be released as a song of a virtual singer (e.g., the debut song of the virtual singer). Note that as the vocal singing voice of the professional singer, vocal singing voice generated by conventional singing voice generation software or the like may be used instead of the singing voice actually sung by the professional singer. In this case, for example, it is possible to create vocal singing voice with ranges or singing techniques that even a professional singer cannot sing as a human.

[0105] For the purpose of creating a virtual singer voice quality conversion model, singers A, B, C, D .. use the microphone 40 to record and store the singing voices (samples) of these singers.

[0106] Figure 15 is a diagram showing an example of the software configuration of the song creation assistance server according to this embodiment. The song creation assistance server 80 includes the singing voice recording unit 101, a virtual voice quality conversion model creation unit 102-2, the input reception unit 103, the voice quality conversion unit 104, the singing voice

editing unit 105, the mix editing unit 106, the completed song output unit 107, and the storage unit 109 (not shown) as main functional units. The software configuration example of the song creation assistance server shown in Figure 15 is different from that in Figure 3 in that it includes the virtual voice quality conversion model creation unit 102-2 and a virtual singer voice quality conversion model 109b-2. In addition, the new singer information DB 109c is replaced with a virtual singer information DB 109c-2.

[0107] How to create the virtual singer voice quality conversion model 109b-2 of a non-existent virtual singer is, for example, as follows:

- The song creation assistance server 10 (the singing voice recording unit 101) records (obtains) singing voices of a plurality of real singers A, B, C, D .. from which a voice quality conversion model is created through the microphone 40. After that, the song creation assistance server 10 (the virtual voice quality conversion model creation unit 102-2) creates a voice quality conversion model for each of the singers based on features of the recorded singing voices of the plurality of singers. Then, by merging (mixing) the voice quality conversion models of the plurality of singers at a predetermined ratio (e.g., 4:3:2:1) to create the virtual singer voice quality conversion model 109b-2, a non-existent voice quality conversion model can be created.

[0108] • The song creation assistance server 10 (the singing voice recording unit 101) records (obtains) singing voices of a plurality of real singers A, B, C, D .. from which a voice quality conversion model is created through the microphone 40. After that, the song creation assistance server 10 (the virtual voice quality conversion model creation unit 102-2) can create the virtual singer voice quality conversion model 109b-2 by mixing and learning the recorded singing voice data of the plurality of singers.

[0109] • In addition, instead of the singing voices of the above real singers A, B, C, D .., it is also possible to use synthesized singing voices of a plurality of virtual singers A', B', C', D' .. generated by conventional singing voice generation software and the like.

[0110] • In addition, the singing voices of real singers may be mixed with synthesized singing voices generated by conventional singing voice generation software or the like to create a voice quality conversion model as in the above three examples.

[0111] After the creation of the virtual singer voice quality conversion model 109b-2, a new singer according to the embodiment 1 can be read as a virtual singer in the song creation assistance system 300 according to the embodiment 3 to implement S2-S33 in the flowcharts shown

in Figures 6-8 in the same way, so the repetitive description is omitted.

[0112] According to this embodiment, completed song data is produced by converting the singing voice obtained by singing the song with the singing ability of a professional singer into the voice quality of a non-existent virtual singer and using the result as the vocal voice. It is possible to debut a non-existent singer in a short period of time at a low cost with a professional singer-level singing voice. Since the virtual singer who debuts in this way (who can also be said to be an android singer) and their songs do not depend on specific singers, it is possible not only to improve the preferences in listening to songs, but also to lower the hurdle for adjusting right relationship such as copyright and expand their uses.

[0113] As described above, these embodiments make it possible to expand debut opportunities for those who have immature skills such as singing skills or acting skills in one aspect. Note that although the present invention has been described by showing certain specific examples using preferable embodiments of the present invention, it is clear that various amendments and modifications can be made to these specific examples without deviating from the broad spirit and scope of the present invention as defined in the claims. In other words, the details of the specific examples and the accompanying drawings

should not be interpreted as limiting the present invention.

[0114] The voice quality-converted singing voice produced by the present invention may be added to the generation of the pre-trained model, to the training data for the voice quality conversion model, or to singing voices of professional singers. In addition, creating a voice quality conversion model of a non-existent human based on the present invention and using a performance audio creation assistance server makes it possible to debut a non-existent human as a virtual voice actor. In addition, if there is a scene in which an actor or voice actor with immature singing ability sings in a movie, TV drama, or anime, the present invention can create a voice quality conversion model of the actor or voice actor, convert the singing voice of a professional singer into the voice quality of the actor or voice actor, and use the singing voice in video creation.

[Reference Signs List]

[0115]

10, 80 song creation assistance server

20 PC

30, 40 microphone

50 CD writer

70 performance audio creation assistance server

100, 300 song creation assistance system

200 performance audio and video creation assistance
system

101 singing voice recording unit

102 voice quality conversion model creation unit

103 input reception unit

104 voice quality conversion unit

105 singing voice editing unit

106 mix editing unit

107 completed song output unit

108 singing voice synthesis unit

109 storage unit

701 voice recording unit

702 voice quality conversion model creation unit

703 input reception unit

704 voice quality conversion unit

705 performance audio editing unit

706 mix editing unit

707 completed video work output unit

709 storage unit

[Document Name] Claims

[Claim 1]

A song creation assistance program for functioning as:

first singing voice recording means for recording singing voice data of a first singer;

voice quality conversion model creation means for creating a voice quality conversion model of the first singer based on the singing voice data of the first singer;

second singing voice recording means for recording first language singing voice data of a song sung by a second singer in a first language, and recording second language singing voice data of the song sung by the second singer or a third singer in a second language; and

voice quality conversion means for converting each of the first language singing voice data and the second language singing voice data into voice quality of the first singer by using the voice quality conversion model of the first singer.

[Claim 2]

The song creation assistance program according to claim 1 for function as:

editing means for mixing the first language singing voice data and the second language singing voice data converted by the voice quality conversion means into the

voice quality of the first singer with melody or accompaniment data; and

output means for outputting the first language singing voice data and the second language singing voice data mixed by the editing means as first language song data sung by the first singer in the first language and second language song data sung by the first singer in the second language.

[Claim 3]

The song creation assistance program according to claim 2, wherein

the output means includes a storage device that stores the first language singing voice data and the second language singing voice data in an external storage medium or a communication device that distributes the first language singing voice data and the second language singing voice data to an external server.

[Claim 4]

The song creation assistance program according to claim 1, wherein

the song is an anime song.

[Claim 5]

The song creation assistance program according to claim 1, wherein

the first singing voice recording means records a plurality of pieces of singing voice data, and

the voice quality conversion model creation means creates a voice quality conversion model of the first singer who does not exist based on the plurality of pieces of singing voice data.

[Claim 6]

A musical reproduction device comprising:

a storage unit that stores the first language song data and the second language song data output by the song creation assistance program according to claim 2;

a first language playback operation unit that is associated with the first language song data;

a second language playback operation unit that is associated with the second language song data; and

a music output unit that plays the first language song when the first language playback operation unit is operated, and plays the second language song when the second language playback operation unit is operated.

[Claim 7]

A song creation assistance device comprising:

first singing voice recording means for recording singing voice data of a first singer;

voice quality conversion model creation means for creating a voice quality conversion model of the first singer based on singing voice of the first singer;

second singing voice recording means for recording first language singing voice data of a song sung by a second singer in a first language, and recording second

language singing voice data of the song sung by the second singer or a third singer in a second language; and

voice quality conversion means for converting each of the first language singing voice data and the second language singing voice data into voice quality of the first singer by using the voice quality conversion model of the first singer.

[Claim 8]

A song creation assistance method, wherein a song creation assistance device executes:

a first singing voice recording step of recording singing voice data of a first singer;

a voice quality conversion model creation step of creating a voice quality conversion model of the first singer based on singing voice of the first singer;

a second singing voice recording step of recording first language singing voice data of a song sung by a second singer in a first language, and recording second language singing voice data of the song sung by the second singer or a third singer in a second language; and

a voice quality conversion step of converting each of the first language singing voice data and the second language singing voice data into voice quality of the first singer by using the voice quality conversion model of the first singer.