



(11)

EP 4 819 123 A1

(12)

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

(43) Date of publication:
07.10.2026 Bulletin 2026/41

(51) International Patent Classification (IPC):
G10G 3/04 ^(2006.01) **G10L 25/51** ^(2013.01)

(21) Application number: **23960227.9**

(52) Cooperative Patent Classification (CPC):
G10G 3/04; G10L 25/51

(22) Date of filing: **01.12.2023**

(86) International application number:
PCT/JP2023/043088

(87) International publication number:
WO 2025/115210 (05.06.2025 Gazette 2025/23)

(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

- **OKADA, Yusuke**
Kawagoe-shi, Saitama 350-8555 (JP)
- **ISHIGURO, Rui**
Kawagoe-shi, Saitama 350-8555 (JP)
- **SUGIURA, Takuma**
Kawagoe-shi, Saitama 350-8555 (JP)

(71) Applicant: **Pioneer Corporation**
Tokyo 113-0021 (JP)

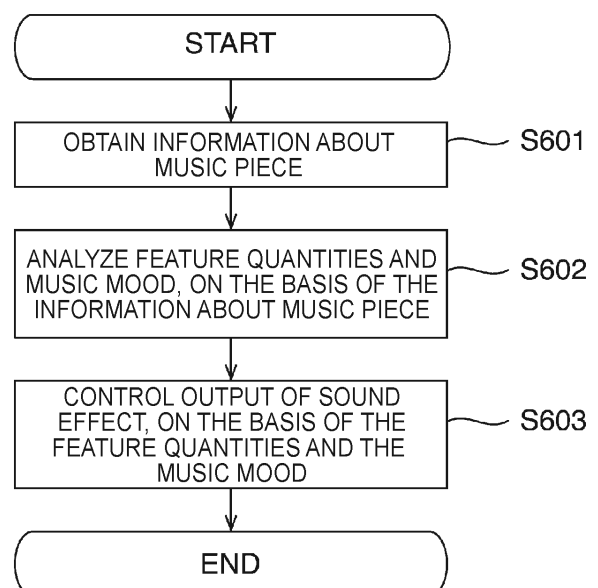
(74) Representative: **Grünecker Patent- und
Rechtsanwälte
PartG mbB
Leopoldstraße 4
80802 München (DE)**

(72) Inventors:
• **TAMURA, Fumio**
Kawagoe-shi, Saitama 350-8555 (JP)

(54) **INFORMATION PROCESSING DEVICE**

(57) To mix sound effects into a music piece in a manner that is more natural for the listener. A beat position of a music piece and a confidence level of the beat position are analyzed, and a sound effect is output at the analyzed beat position. An output level of the sound effect is gradually decreased to zero after a beat position at which a value of the confidence level becomes equal to or less than a predetermined confidence level value.

FIG. 6



EP 4 819 123 A1

Description

Technical Field

[0001] The present invention relates to an information processing apparatus. 5

Background

[0002] A technique is known in which sound effects are mixed into a music piece so that a listener can experience the atmosphere of a live venue. For example, Patent Documents 1 to 4 disclose techniques for mixing sound effects into a music piece in a manner that is natural for the listener. 10 15

Related Art

Patent Document

[0003]

Patent Document 1: JP2021162707 A
 Patent Document 2: WO2023/054236
 Patent Document 3: WO2023/054237
 Patent Document 4: JP2023050570 A 20 25

Summary of the Invention

Problem to be Solved by the Invention

[0004] In recent years, streaming playback has been widely used as a method for playing a music piece. To mix sound effects into a music piece being streamed, the sound effect mixing processing needs to be performed in real-time processing. The techniques disclosed in Patent Documents 1 to 4 do not take real-time processing into consideration. 30 35

[0005] An example of a problem to be solved by the present invention is to mix sound effects into a music piece in a manner that is more natural for the listener. 40

Solution to Problem

[0006] In order to solve the above problems, the invention set forth in claim 1 includes: 45

a music piece acquisition processing unit configured to acquire information relating to audio of a music piece; 50
 a music piece feature quantity analysis unit configured to analyze a beat position of the music piece and a confidence level of the beat position, based on information relating to the audio of the music piece acquired by the music piece acquisition processing unit; and 55
 a sound effect output control unit configured to output a sound effect at the beat position analyzed by the

music piece feature quantity analysis unit, wherein the sound effect output control unit is configured to, after a beat position at which a value of the confidence level becomes equal to or less than a predetermined confidence level value, gradually decrease an output level of the sound effect to zero.

[0007] The invention set forth in claim 8 is an information processing method executed by a computer, including: 10

a music piece acquisition processing step of acquiring information relating to audio of a music piece; a music piece feature quantity analysis step of analyzing a beat position of the music piece and a confidence level of the beat position, based on information relating to the audio of the music piece acquired in the music piece acquisition processing step; and a sound effect output control step of outputting a sound effect at the beat position analyzed in the music piece feature quantity analysis step, wherein in the sound effect output control step, an output level of the sound effect is gradually decreased to zero after a beat position at which a value of the confidence level becomes equal to or less than a predetermined confidence level value. 15 20 25

[0008] The invention set forth in claim 9 is an information processing program for causing a computer to execute the information processing method according to claim 8. 30

[0009] The invention set forth in claim 10 is a computer-readable storage medium storing the information processing program according to claim 9. 35

Brief Description of the Drawings

[0010]

[Fig. 1] Fig. 1 is a diagram illustrating a sound effect mixing apparatus 100 according to an embodiment of the present invention.

[Fig. 2] Fig. 2 is a diagram illustrating processing in the sound effect mixing apparatus 100.

[Fig. 3] Fig. 3 is a diagram illustrating an example of a music piece input by a music piece input unit 110 and a sound effect output by a sound effect output unit 130.

[Fig. 4] Fig. 4 is a diagram illustrating a control unit 160. 50

[Fig. 5] Fig. 5 is a diagram illustrating that parts of frames overlap.

[Fig. 6] Fig. 6 is a diagram illustrating an example of processing operations in the control unit 160.

[Fig. 7] Fig. 7 is a diagram illustrating an example of a relationship between a recent average value of volume levels of a music piece and a second threshold value. 55

[Fig. 8] Fig. 8 is a diagram illustrating an example of a relationship between a second time and an intensity of a beat of a music piece.

[Fig. 9] Fig. 9 is a diagram illustrating an example of a relationship between a second time and an intensity of a beat of a music piece.

[Fig. 10] Fig. 10 is a diagram illustrating a rising difference.

Modes for Carrying Out the Invention

[0011] An information processing apparatus according to an embodiment of the present invention includes:

a music piece acquisition processing unit configured to acquire information relating to audio of a music piece;

a music piece feature quantity analysis unit configured to analyze a beat position of the music piece and a confidence level of the beat position, based on information relating to the audio of the music piece acquired by the music piece acquisition processing unit; and

a sound effect output control unit configured to output a sound effect at the beat position analyzed by the music piece feature quantity analysis unit, wherein the sound effect output control unit is configured to, after a beat position at which a value of the confidence level becomes equal to or less than a predetermined confidence level value, gradually decrease an output level of the sound effect to zero. The information relating to the audio of the music piece may include an audio signal of the music piece and/or streaming data. The information processing apparatus may execute real-time processing. Accordingly, in the present embodiment, when hand clapping mixed into a music piece deviates from the beat position of the music piece, mixing of the hand clapping can be stopped without causing the listener to feel unnaturalness. As a result, the present embodiment enables the listener to more strongly experience the atmosphere of a live venue.

[0012] The sound effect output control unit may be configured to, after a beat position at which the value of the confidence level becomes equal to or less than the predetermined confidence level value, output the sound effect at a constant tempo while gradually decreasing the output level of the sound effect to zero. The constant tempo may be determined based on an interval between the beat position at which the value of the confidence level becomes equal to or less than the predetermined confidence level value and a beat position immediately preceding the beat position. The constant tempo may be determined based on an average value of intervals between two adjacent beat positions obtained until the value of the confidence level becomes equal to or less than the predetermined confidence level value. Accord-

ingly, when hand clapping mixed into a music piece deviates from the beat position of the music piece, mixing of the hand clapping can be stopped more reliably without causing the listener to feel unnaturalness.

[0013] A time from when the output level of the sound effect starts to gradually decrease to when the output level reaches zero may be equal to or less than a predetermined time. Accordingly, when hand clapping mixed into a music piece deviates from the beat position of the music piece, mixing of the hand clapping can be stopped rapidly.

[0014] An information processing method according to an embodiment of the present invention is an information processing method executed by a computer, comprising:

a music piece acquisition processing step of acquiring information relating to audio of a music piece; a music piece feature quantity analysis step of analyzing a beat position of the music piece and a confidence level of the beat position, based on information relating to the audio of the music piece acquired in the music piece acquisition processing step; and a sound effect output control step of outputting a sound effect at the beat position analyzed in the music piece feature quantity analysis step, wherein in the sound effect output control step, an output level of the sound effect is gradually decreased to zero after a beat position at which a value of the confidence level becomes equal to or less than a predetermined confidence level value. Accordingly, in the present embodiment, when hand clapping mixed into a music piece deviates from the beat position of the music piece, mixing of the hand clapping can be stopped without causing the listener to feel unnaturalness. As a result, the present embodiment enables the listener to more strongly experience the atmosphere of a live venue.

[0015] An information processing program according to an embodiment of the present invention causes a computer to execute the above information processing method. Accordingly, the present embodiment enables the listener to more strongly experience the atmosphere of a live venue.

[0016] A storage medium according to an embodiment of the present invention stores the above information processing program. Accordingly, in the present embodiment, the above information processing program can be distributed independently in addition to being incorporated into a device, thereby enabling easy version upgrades and the like.

Embodiment

[0017] <Sound effect mixing apparatus 100>

[0017] Fig. 1 is a diagram illustrating a sound effect mixing apparatus 100 according to an embodiment of the

present invention. The sound effect mixing apparatus 100 includes a music piece input unit 110, a storage unit 120, a sound effect output unit 130, a mixing unit 140, an external output unit 150, and a control unit 160.

[0018] In the sound effect mixing apparatus 100 according to the present embodiment, for example, as shown in Fig. 2, a sound effect output from the sound effect output unit 130 is mixed by the mixing unit 140 with a music piece input from outside to the music piece input unit 110, and the music piece mixed with the sound effect is output to the outside from the external output unit 150. At this time, in the sound effect mixing apparatus 100 according to the present embodiment, the storage unit 120 stores sound source data for sound effects, and the control unit 160 controls output of the sound effect from the sound effect output unit 130 by real-time processing.

[0019] The music piece input unit 110 receives, from outside, information relating to the audio of a music piece, and outputs the music piece (for example, an audio signal of the music piece) to the mixing unit 140. The information relating to the audio of the music piece includes, for example, an audio signal of the music piece and streaming data. When the information relating to the audio of the music piece is streaming data of the music piece, the music piece input unit 110 converts the streaming data of the music piece into an audio signal of the music piece and outputs it to the mixing unit 140. The music piece input unit 110 includes, for example, signal input terminals that receive an input of the audio signal of the music piece from a playback apparatus that plays the music piece, and a communication apparatus that receives, from outside, an input of the streaming data of the music piece. The music piece input unit 110 may also include an audio input apparatus (for example, a microphone) that receives an input of audio of the music piece.

[0020] The storage unit 120 is a storage apparatus that stores information, such as a hard disk drive, a solid-state drive, or a memory. The storage unit 120 stores sound source data for sound effects.

[0021] The sound effect output unit 130 outputs a sound effect (for example, an audio signal of the sound effect) to the mixing unit 140. The sound effect output unit 130 acquires, for example, sound source data for sound effects stored in the storage unit 120, generates an audio signal of the sound effect from the acquired sound source data, and outputs the generated audio signal of the sound effect to the mixing unit 140.

[0022] Fig. 3 is a diagram illustrating an example of a music piece input to the music piece input unit 110 and a sound effect output from the sound effect output unit 130. In the example illustrated in Fig. 3, a first sound effect (such as cheering or applause) is output during the music piece and at a beginning portion and an ending portion of the music piece. A second sound effect (such as ambient sound) is continuously output from before the start of the music piece to after the end thereof. A third sound effect (such as hand clapping) is output during the music piece in synchronization with the beat or tempo of the music

piece.

[0023] The mixing unit 140 mixes an audio signal of a music piece output from the music piece input unit 110 with an audio signal of a sound effect output from the sound effect output unit 130, and outputs an audio signal of the music piece mixed with the sound effect. The mixing unit 140 is, for example, an apparatus that adds a plurality of signals and outputs the added signal. The mixing unit 140 adds the audio signal of the music piece output from the music piece input unit 110 and the audio signal of the sound effect output from the sound effect output unit 130, and outputs the added signal.

[0024] The external output unit 150 outputs, to the outside, the music piece that has been output from the mixing unit 140 (i.e., the music piece to which sound effects have been mixed). The external output unit 150 includes, for example, an audio output apparatus (e.g., a speaker) that outputs audio, signal output terminals that output an audio signal, and a communication apparatus that transmits information to another apparatus.

[0025] The control unit 160 controls output of the sound effect from the sound effect output unit 130. The control unit 160 is an information processing apparatus such as a computer. The control unit 160 of the sound effect mixing apparatus 100 according to the present embodiment performs real-time processing. Specifically, the control unit 160 performs real-time processing on information relating to the audio of a music piece (e.g., an audio signal or streaming data) input from the outside by the music piece input unit 110, and controls output of the sound effect from the sound effect output unit 130, based on feature quantities of the music piece obtained by the processing.

[0026] Fig. 4 is a diagram illustrating the control unit 160. The control unit 160 includes a music piece acquisition processing unit 161, a music piece feature quantity analysis unit 162, and a sound effect output control unit 163.

[0027] The music piece acquisition processing unit 161 acquires information relating to the audio of a music piece (e.g., an audio signal or streaming data) input from the outside via the music piece input unit 110. In this case, for example, the music piece acquisition processing unit 161 sequentially acquires the information relating to the audio of the music piece each time such information (e.g., an audio signal or streaming data) is input to the music piece input unit 110. In other words, the music piece acquisition processing unit 161 acquires the audio signal or streaming data of the music piece input from the outside via the music piece input unit 110 in real time.

[0028] The music piece feature quantity analysis unit 162 analyzes feature quantities of the music piece, based on information relating to the audio of the music piece (e.g., the audio signal or streaming data of the music piece) acquired by the music piece acquisition processing unit 161. The feature quantities of the music piece include, for example, the volume level of the music piece, the position of beats in the music piece, the tempo of the

music piece (e.g., beats per minute, BPM), and the intensity of the beats in the music piece.

[0029] The music piece feature quantity analysis unit 162 may, for example, analyze the feature quantities of the music piece by frame processing. In this case, as shown in Fig. 5, it is preferable that the frames partially overlap. Accordingly, the feature quantities of the music piece can be analyzed without being affected by small fluctuations in the audio signal. When the music piece feature quantity analysis unit 162 performs frame processing, the volume level of the music piece in each frame is, for example, the RMS (Root Mean Square) value in that frame.

[0030] The sound effect output control unit 163 controls output of the sound effect from the sound effect output unit 130, based on the feature quantities of the music piece analyzed by the music piece feature quantity analysis unit 162.

[0031] The sound effect output control unit 163 controls the volume level of the sound effect output from the sound effect output unit 130, based, for example, on the volume level of the music piece analyzed by the music piece feature quantity analysis unit 162. Accordingly, the volume of the mixed sound effect can be prevented from becoming excessively high or too low relative to the volume of the music piece. This enables a more natural addition of the sound effect to the music piece for the listener, allowing the listener to better experience the atmosphere of a live venue.

[0032] Additionally, as shown in Fig. 4, the control unit 160 may further include a music mood analysis unit 164. The music mood analysis unit 164 analyzes the music mood of the music piece, based on information relating to the audio of the music piece (e.g., the audio signal or streaming data of the music piece) acquired by the music piece acquisition processing unit 161 and on the feature quantities of the music piece acquired by the music piece feature quantity analysis unit 162. The sound effect output control unit 163 may control output of the sound effect from the sound effect output unit 130, based on the music mood analyzed by the music mood analysis unit 164. For example, the sound effect output control unit 163 may control the volume level of the sound effect output from the sound effect output unit 130. In this case, the storage unit 120 may store sound source data for the sound effect prepared for each volume level and music mood of the music piece, and the sound effect output control unit 163 may determine the sound source data for the sound effect to be output from the sound effect output unit 130, based on the analyzed volume level and music mood of the music piece.

[0033] In addition, the control unit 160 may further include a mode selection unit 165, as shown in Fig. 4. The mode selection unit 165 selects one mode from a plurality of modes relating to, for example, the music mood or genre of a music piece, and the scale of a live venue. At this time, the mode selection unit 165 may be configured to select one mode from the plurality of

modes, based on, for example, the analyzed feature quantities or music mood of the music piece. The sound effect output control unit 163 may then control the output of the sound effect that is output from the sound effect output unit 130, based on the mode input by the mode selection unit 165.

[0034] For example, the plurality of modes may include modes corresponding to the scale of a live venue. The plurality of modes may include, for example, a mode for large-scale venues such as stadiums, outdoor festivals, and arenas; a mode for medium-scale venues such as halls and medium-to-large live houses; and a mode for small-scale venues such as small live houses and music bars. In this case, for example, the storage unit 120 may be configured to store sound effects for large-scale venues, sound effects for medium-scale venues, and sound effects for small-scale venues.

[0035] Further, as shown in Fig. 1, the sound effect mixing apparatus 100 may further include a user input unit 170, and the mode selection unit 165 may be configured to select one mode from the plurality of modes, based on input from the user input unit 170. The user input unit 170 is an input apparatus that receives input of information from a user, such as a button, a keyboard, a touch panel, a camera, or a microphone.

<Processing operations in the control unit 160>

[0036] Fig. 6 is a diagram illustrating an example of processing operations performed by the control unit 160. The processing operations shown in Fig. 6 are executed, for example, at predetermined time intervals. When the music piece feature quantity analysis unit 162 performs frame processing, the processing operations shown in Fig. 6 are executed, for example, for each frame. The music piece acquisition processing unit 161 acquires information relating to the audio of a music piece (for example, an audio signal or streaming data of the music piece) input from outside to the music piece input unit 110 (step S601). The music piece feature quantity analysis unit 162 and the music mood analysis unit 164 analyze the feature quantities and the music mood of the music piece, based on the information relating to the audio of the music piece acquired by the music piece acquisition processing unit 161 (step S602). The sound effect output control unit 163 controls the output of a sound effect from the sound effect output unit 130, based on the feature quantities and the music mood of the music piece analyzed by the music piece feature quantity analysis unit 162 and the music mood analysis unit 164 (step S603).

<Output of first sound effect (such as cheers or applause) at beginning of music piece>

[0037] In a live venue, when a music piece is performed, cheers or applause from an audience may occur at the beginning of the performance. When the sound effect mixing apparatus 100 receives only information

relating to the audio of the music piece (such as an audio signal or streaming data) from a playback apparatus and does not receive information relating to a start time of the music piece, the sound effect mixing apparatus 100 cannot mix cheers or applause into the beginning portion of the music piece, based on the start time of the music piece. Further, the magnitude and manner of such cheers or applause depend on the music mood of the music piece performed in the live venue.

[0038] Accordingly, in the present embodiment, the music mood analysis unit 164 analyzes the music mood of the music piece, based on information relating to the audio of the music piece obtained in a period from when the volume level of the music piece exceeds a predetermined threshold value (a first threshold value) to when a predetermined time (a first time) has elapsed since. The sound effect output control unit 163 outputs a first sound effect, based on the music mood analyzed by the music mood analysis unit 164 when the first time has elapsed since the volume level of the music piece exceeded the first threshold value.

[0039] Accordingly, in the present embodiment, even when the sound effect mixing apparatus 100 does not receive input of information relating to a start time of a music piece, cheers or applause can be mixed into a beginning portion of the music piece. Further, in the present embodiment, the music mood of the music piece is analyzed using only information relating to the audio of the music piece up to a timing at which a first sound effect (such as cheers or applause) is output, that is, a timing at which a first time has elapsed since the volume level of the music piece exceeded a first threshold value. Accordingly, in the present embodiment, cheers or applause can be mixed into the music piece by real-time processing.

[0040] Here, the first time is related to, for example, a time required for an audience at a live venue to perceive the start of a performance. In general, it takes about 1 second to 1.4 seconds from the start of the performance for the audience at the live venue to perceive the start of the performance. Accordingly, the first time is preferably set to 1 second or more and 1.4 seconds or less.

[0041] Further, the time required for an audience at a live venue to perceive the start of a performance varies depending on the size of the live venue. The smaller the live venue, the shorter the distance between the audience and the performer, and the shorter the time required for the audience to perceive the start of the performance. Accordingly, the plurality of modes may include modes corresponding to the scale of the live venue, and the first time may be determined based on the mode selected by the mode selection unit 165. In this case, the first time is preferably determined so as to increase as the scale of the live venue increases. For example, when a mode for a small-scale venue is selected, the first time may be set to 1 second; when a mode for a medium-scale venue is selected, the first time may be set to 1.2 seconds; and when a mode for a large-scale venue is selected, the first time may be set to 1.4 seconds.

[0042] After output of a first sound effect (such as cheers or applause) has started, the music mood of the music piece may change. Accordingly, the music mood analysis unit 164 may analyze the music mood of the music piece, based on information relating to the audio of the music piece after a first time has elapsed since the volume level of the music piece exceeded a first threshold value. The sound effect output control unit 163 may change the sound effect to be output based on a change in the music mood when the music mood analyzed by the music mood analysis unit 164 has changed.

[0043] For example, when the music mood of the music piece changes from a calm music mood to an intense music mood, the sound effect output control unit 163 may be configured to output a sound effect obtained by mixing a sound effect for the intense music mood with a sound effect for the calm music mood.

[0044] Further, when the music mood of the music piece changes from a first music mood to a second music mood, the sound effect output control unit 163 may be configured to change the sound effect to be output from a sound effect for the first music mood to a sound effect for the second music mood. At the time of this change from the sound effect for the first music mood to the sound effect for the second music mood, the sound effect output control unit 163 may gradually decrease the output level of the sound effect for the first music mood to zero (fade-out), and gradually increase the output level of the sound effect for the second music mood from zero (fade-in).

<Output of first sound effect (such as cheers or applause) at end of music piece>

[0045] In a live venue, when a music piece is performed, cheers or applause from an audience may occur at an ending portion of the performance. When the sound effect mixing apparatus 100 receives only information relating to the audio of the music piece (such as an audio signal or streaming data) from a playback apparatus and does not receive information relating to an end time of the music piece, the sound effect mixing apparatus 100 cannot mix cheers or applause into an ending portion of the music piece, based on the end time of the music piece.

[0046] Accordingly, in the present embodiment, the sound effect output control unit 163 outputs a first sound effect (such as cheers or applause) from the sound effect output unit 130 when the volume level of the music piece is equal to or less than a predetermined threshold value (a second threshold value).

[0047] At this time, the timing at which the audience perceives that the performance has ended varies depending on the volume level of the music piece. Accordingly, in the present embodiment, the second threshold value is determined based on a value relating to the volume level of the music piece. The value relating to the volume level of the music piece is, for example, a recent average value of the volume level of the music

piece. The recent average value of the volume level of the music piece may be a recent moving average value of the volume level of the music piece or an average value obtained using a time-constant circuit. When the music piece feature quantity analysis unit 162 performs frame processing, the recent moving average value of the volume level of the music piece may be, for example, an average value of RMS values in a predetermined number (two or more) of most recent frames.

[0048] When the second threshold value is determined based on a recent average value of the volume level of the music piece, the second threshold value may be set, as shown in Fig. 7, to a value obtained by subtracting an offset value from the recent average value of the volume level of the music piece. The value obtained by subtracting the offset value from the recent average value of the volume level of the music piece may be negative. Accordingly, a minimum value (a minimum threshold value) may be set for the second threshold value. As shown in Fig. 7, when the value obtained by subtracting the offset value from the recent average value of the volume level of the music piece is equal to or greater than the minimum threshold value, the second threshold value may be set to the value obtained by subtracting the offset value from the recent average value of the volume level of the music piece; and when the value obtained by subtracting the offset value from the recent average value of the volume level of the music piece is less than the minimum threshold value, the second threshold value may be set to the minimum threshold value. In Fig. 7, a thin solid line represents the recent average value of the volume level of the music piece, a thick solid line represents the second threshold value, a one-dot chain line represents the minimum threshold value, and a dashed line represents the value obtained by subtracting the offset value from the recent average value of the volume level of the music piece.

[0049] Determination of the end of the music piece (i.e., determination of whether the volume level of the music piece is equal to or less than the second threshold value) may be performed in an ending portion of the music piece. Accordingly, the sound effect output control unit 163 may be configured not to perform determination of whether the volume level of the music piece is equal to or less than the second threshold value during a predetermined period (a first period) from the start of the music piece. Here, the start of the music piece may be, for example, a timing at which input of information relating to the audio of the music piece is started, or a timing at which the volume level of the music piece exceeds a first threshold value. Further, as will be described in detail below, when information relating to the audio of the music piece (such as an audio signal or streaming data) includes two or more music pieces, the start of a subsequent music piece among two consecutive music pieces is a boundary between a preceding music piece and the subsequent music piece.

<Determination of ending portion of music piece>

[0050] Depending on the music piece, there may be a silent portion (a break). When a portion in which the volume level is equal to or less than the second threshold value is determined to be an ending portion of the music piece as described above, cheers or applause corresponding to the ending portion of the music piece would also be mixed into such a break portion.

[0051] Accordingly, in the present embodiment, when the volume level of the music piece becomes equal to or less than the second threshold value and a state in which the volume level of the music piece is equal to or less than the second threshold value continues for a predetermined time (a second time), the sound effect output control unit 163 outputs a first sound effect (such as cheers or applause) from the sound effect output unit 130 when the second time has elapsed since the volume level of the music piece became equal to or less than the second threshold value.

[0052] In this case, the second time may be a constant. The second time may be set to an integral multiple of a beat of the music piece, or may be set to a value obtained by adding a predetermined time (for example, 100 msec) to an integral multiple of the beat of the music piece.

[0053] Further, a music piece having a stronger beat tends to have a longer break period. Accordingly, the second time may be determined based on a value relating to the intensity of the beat of the music piece. In this case, the second time may be set to increase as the intensity of the beat of the music piece increases.

[0054] In this case, the second time may be configured to vary linearly with respect to the intensity of the beat of the music piece, as shown in Fig. 8, or may be configured to vary stepwise with respect to the intensity of the beat of the music piece, as shown in Fig. 9. When the second time varies stepwise with respect to the intensity of the beat of the music piece (Fig. 9), the second time at each step may be set to an integral multiple of a beat of the music piece, or may be set to a value obtained by adding a predetermined time (for example, 100 msec) to an integral multiple of the beat of the music piece.

[0055] The intensity of the beat of the music piece is determined, for example, based on the volume level of the music piece. When the music piece feature quantity analysis unit 162 performs frame processing and the volume level of the music piece is an RMS value for each frame, the intensity of the beat of the music piece may be, for example, a maximum value among rising differences in a predetermined number (two or more) of most recent frames. Here, the rising difference is a difference between an RMS value of a later frame and an RMS value of a preceding frame when the RMS value of the later frame is greater than the RMS value of the preceding frame in two consecutive frames (Fig. 10).

<Output of third sound effect (such as hand clapping)>

[0056] In a live venue, an audience may keep rhythm with a music piece by hand clapping or foot stamping. Accordingly, in the present embodiment, the music piece feature quantity analysis unit 162 analyzes beat positions of the music piece, based on information relating to the audio of the music piece acquired by the music piece acquisition processing unit 161. The sound effect output control unit 163 outputs a third sound effect (such as hand clapping) at the beat positions analyzed by the music piece feature quantity analysis unit 162.

[0057] In a playback apparatus that plays back a music piece, operations such as pause and fast-forward may be performed. When the sound effect mixing apparatus 100 receives only information relating to the audio of the music piece (such as an audio signal or streaming data) from the playback apparatus and does not receive information relating to operations performed on the playback apparatus, hand clapping mixed into the music piece may become significantly misaligned with beat positions of the music piece after a timing at which such operations (e.g., pause or fast-forward) are performed.

[0058] Accordingly, in the present embodiment, the music piece feature quantity analysis unit 162 analyzes not only beat positions of the music piece but also a confidence level of each beat position, based on information relating to the audio of the music piece acquired by the music piece acquisition processing unit 161. The confidence level of a beat position is an indicator representing accuracy of the beat position. The sound effect output control unit 163 gradually decreases an output level of a third sound effect to zero from a beat position at which the confidence level becomes equal to or less than a predetermined confidence level value.

[0059] At this time, a time from when the output level of the sound effect starts to gradually decrease to when the output level becomes zero is preferably set to be equal to or less than a predetermined time (for example, 10 seconds). With this configuration, when hand clapping mixed into the music piece becomes misaligned with beat positions of the music piece, mixing of the hand clapping can be stopped without causing a listener to feel unnaturalness.

[0060] Further, in this case, the sound effect output control unit 163 may be configured to output the third sound effect at a constant tempo while gradually decreasing an output level of the third sound effect to zero, from a beat position at which the confidence level becomes equal to or less than a first confidence level value. Here, the constant tempo is preferably determined based on an interval between the beat position at which the confidence level becomes equal to or less than the first confidence level value and a beat position immediately preceding the beat position. Alternatively, the constant tempo may be determined based on an average value of intervals between adjacent two beat positions up to a point at which the confidence level becomes equal to or

less than a predetermined confidence level value. With this configuration, when hand clapping mixed into the music piece becomes misaligned with beat positions of the music piece, mixing of the hand clapping can be stopped more reliably without causing a listener to feel unnaturalness.

[0061] Analysis of the beat positions of the music piece and the confidence levels of the beat positions may be performed, for example, using a technique described in Masataka Goto and Yoichi Muraoka, "A Real-time Beat Tracking System for Audio Signals (Beat Tracking for Music without Drum-sounds)", IEICE Transactions D, February 1998, Vol. J81-D2, pp. 227-237.

<Determination of a boundary between two music pieces>

[0062] Analysis of feature quantities of a music piece becomes more accurate as the amount of information that can be used increases. Accordingly, when information relating to the audio of a music piece (such as an audio signal or streaming data) includes two or more music pieces, it is preferable to perform analysis of feature quantities of a subsequent music piece starting from a boundary between a preceding music piece and the subsequent music piece. In this case, if a timing at which a volume level of the subsequent music piece exceeds a first threshold value is regarded as the boundary between the preceding music piece and the subsequent music piece, information of the subsequent music piece during a period in which the volume level is less than the first threshold value is not used for the analysis of feature quantities of the subsequent music piece. Further, if a timing at which a volume level of the preceding music piece becomes equal to or less than a second threshold value is regarded as the boundary between the preceding music piece and the subsequent music piece, information of the preceding music piece during a period in which the volume level is less than the second threshold value may be used for the analysis of feature quantities of the subsequent music piece.

[0063] Accordingly, when information relating to the audio of a music piece (such as an audio signal or streaming data) includes two or more music pieces, for example, a third threshold value smaller than the second threshold value may be set, and a period during which a volume level of a preceding music piece is equal to or less than the third threshold value and continues for a predetermined period (a second period) may be regarded as a silent portion between the preceding music piece and a subsequent music piece. In this case, when information relating to the audio of a music piece (such as an audio signal or streaming data) includes two or more music pieces, the music piece feature quantity analysis unit 162 and the music mood analysis unit 164 determine, when a state in which the volume level of the preceding music piece is equal to or less than the third threshold value continues for the second period, a timing relating to a

period during which the volume level of the preceding music piece remains equal to or less than the third threshold value (for example, a timing at which the volume level of the preceding music piece becomes equal to or less than the third threshold value, or a timing after the second period has elapsed from the timing at which the volume level of the preceding music piece becomes equal to or less than the third threshold value) as a boundary between the preceding music piece and the subsequent music piece (i.e., a start of the subsequent music piece), and analyze feature quantities and a music mood of the subsequent music piece based on information relating to the audio of the music piece after the boundary between the preceding music piece and the subsequent music piece.

[0064] In particular, when information relating to the audio of a music piece (such as an audio signal or streaming data) includes two or more music pieces and the sound effect output control unit 163 outputs a first sound effect (such as cheers or applause) at a timing when a first time has elapsed from a time at which a volume level of a subsequent music piece exceeds a first threshold value, the music mood analysis unit 164 may analyze a music mood of the subsequent music piece, based on information relating to the audio of the music piece during a period between a boundary between a preceding music piece and the subsequent music piece and the timing when the first time has elapsed from the time at which the volume level of the subsequent music piece exceeds the first threshold value, and the sound effect output control unit 163 may output the first sound effect based on the analyzed music mood.

[0065] A silent portion between two consecutive music pieces may include noise. Accordingly, the sound effect output control unit 163 may be configured to determine that a state in which the volume level of the preceding music piece is equal to or less than the third threshold value continues, even if a momentary increase in the volume level (noise) occurs after the volume level of the preceding music piece has fallen to or below the third threshold value.

[0066] The present invention has been described above with reference to preferred embodiments. Although the present invention has been described with reference to specific examples herein, various modifications and changes may be made to these examples without departing from the spirit and scope of the present invention as set forth in the claims.

List of Reference Signs

[0067]

100 sound effect mixing apparatus
110 music piece input unit
120 storage unit
130 sound effect output unit
140 mixing unit

150 external output unit
160 control unit
161 music piece acquisition processing unit
162 music piece feature quantity analysis unit
163 sound effect output control unit
164 music mood analysis unit
165 mode selection unit
170 user input unit

10 Claims

1. An information processing apparatus, comprising:

15 a music piece acquisition processing unit configured to acquire information relating to audio of a music piece;
a music piece feature quantity analysis unit configured to analyze a beat position of the music piece and a confidence level of the beat position, based on information relating to the audio of the music piece acquired by the music piece acquisition processing unit; and
20 a sound effect output control unit configured to output a sound effect at the beat position analyzed by the music piece feature quantity analysis unit,
25 wherein the sound effect output control unit is configured to, after a beat position at which a value of the confidence level becomes equal to or less than a predetermined confidence level value, gradually decrease an output level of the sound effect to zero.

2. The information processing apparatus according to claim 1, wherein the sound effect output control unit is configured to, after a beat position at which the value of the confidence level becomes equal to or less than the predetermined confidence level value, output the sound effect at a constant tempo while gradually decreasing the output level of the sound effect to zero.

3. The information processing apparatus according to claim 2, wherein the constant tempo is determined based on an interval between the beat position at which the value of the confidence level becomes equal to or less than the predetermined confidence level value and a beat position immediately preceding the beat position.

4. The information processing apparatus according to claim 2, wherein the constant tempo is determined based on an average value of intervals between two adjacent beat positions obtained until the value of the confidence level becomes equal to or less than the predetermined confidence level value.

5. The information processing apparatus according to

claim 1, wherein a time from when the output level of the sound effect starts to gradually decrease to when the output level reaches zero is equal to or less than a predetermined time.

5

6. The information processing apparatus according to any one of claims 1 to 5, wherein the information relating to the audio of the music piece includes an audio signal of the music piece and/or streaming data.

10

7. The information processing apparatus according to claim 6, wherein the information processing apparatus executes real-time processing.

15

8. An information processing method executed by a computer, comprising:

a music piece acquisition processing step of acquiring information relating to audio of a music piece;

20

a music piece feature quantity analysis step of analyzing a beat position of the music piece and a confidence level of the beat position, based on information relating to the audio of the music piece acquired in the music piece acquisition processing unit; and

25

a sound effect output control step of outputting a sound effect at the beat position analyzed in the music piece feature quantity analysis unit, wherein in the sound effect output control step, an output level of the sound effect is gradually decreased to zero after a beat position at which a value of the confidence level becomes equal to or less than a predetermined confidence level value.

30

35

9. An information processing program for causing a computer to execute the information processing method according to claim 8.

40

10. A computer-readable storage medium storing the information processing program according to claim 9.

45

50

55

FIG. 1

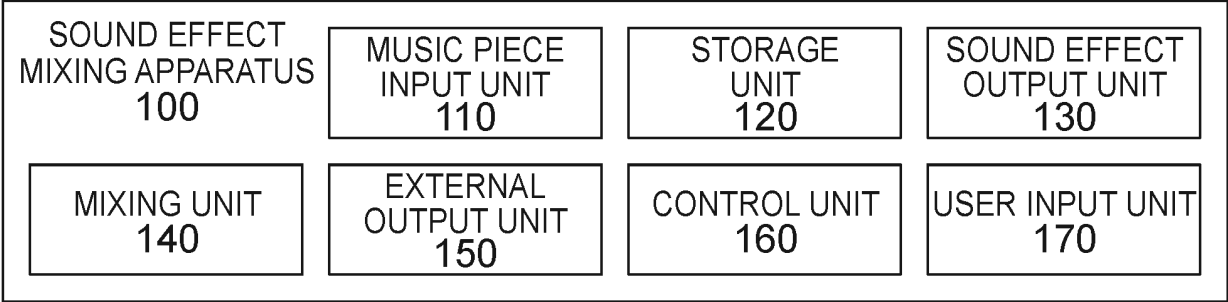


FIG. 2

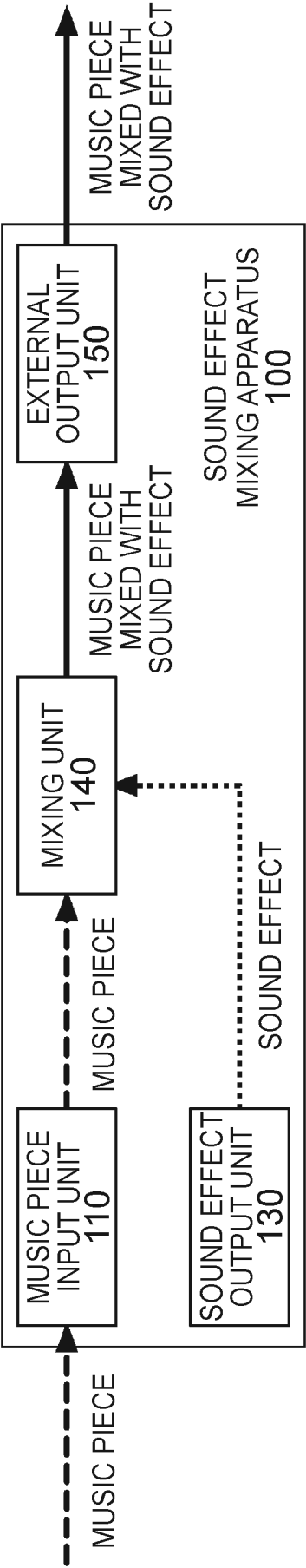


FIG. 3

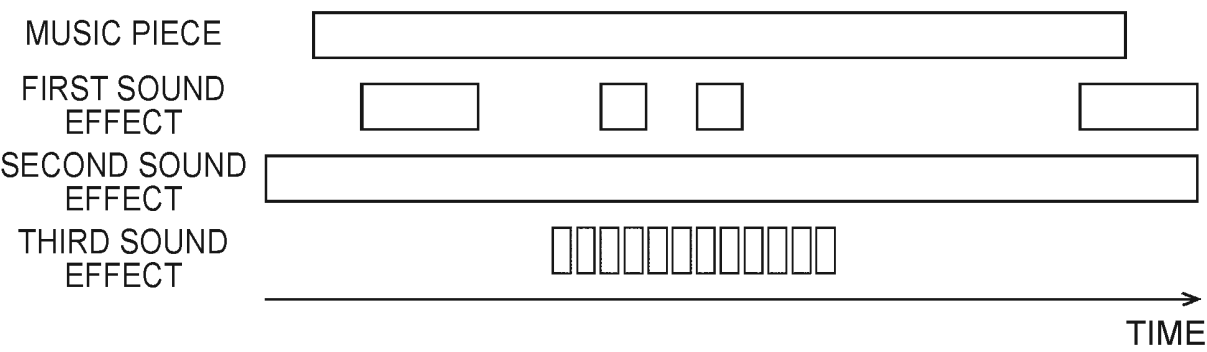


FIG. 4

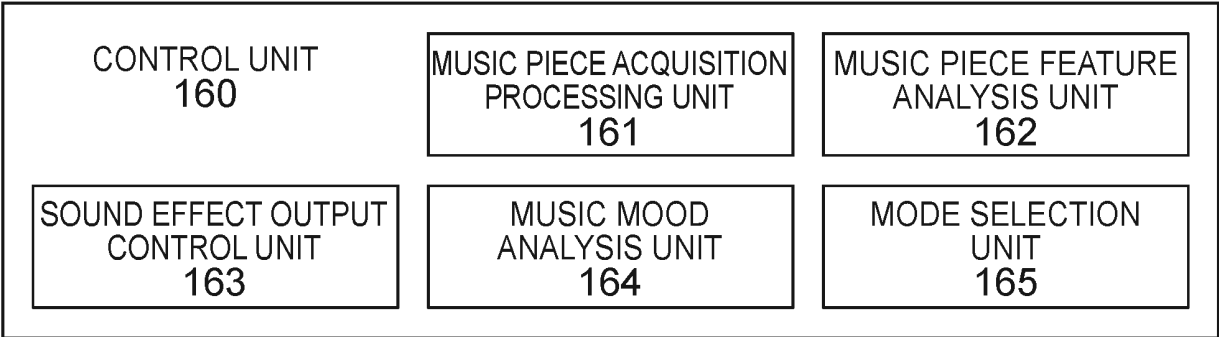


FIG. 5

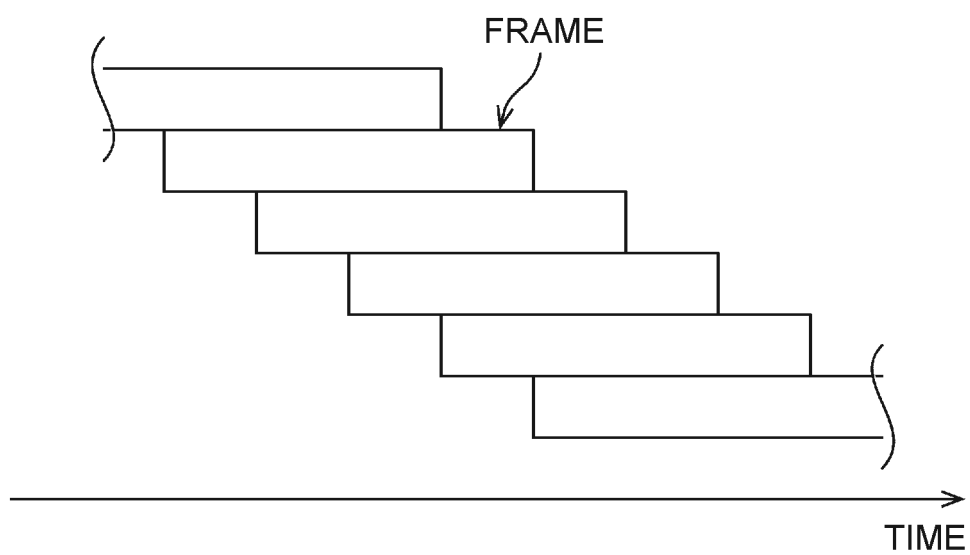


FIG. 6

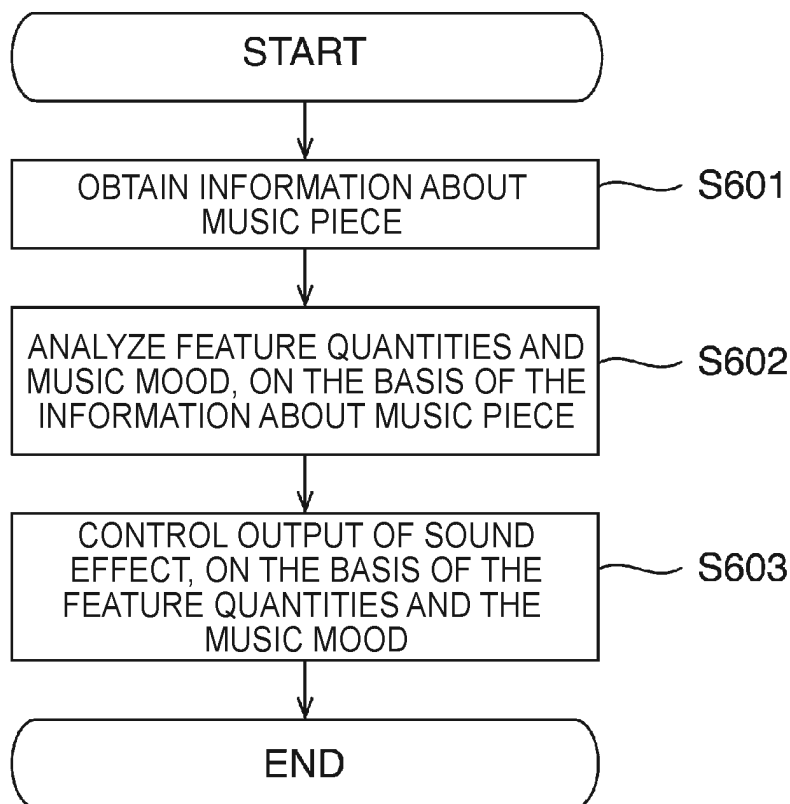


FIG. 7

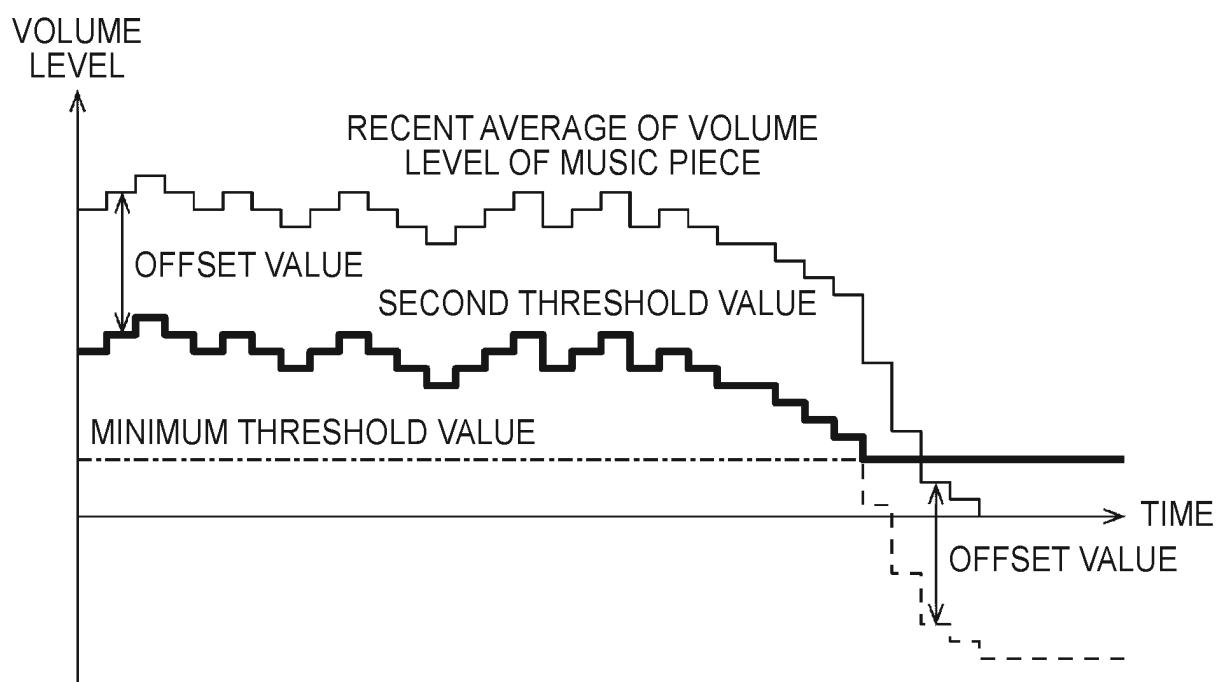


FIG. 8

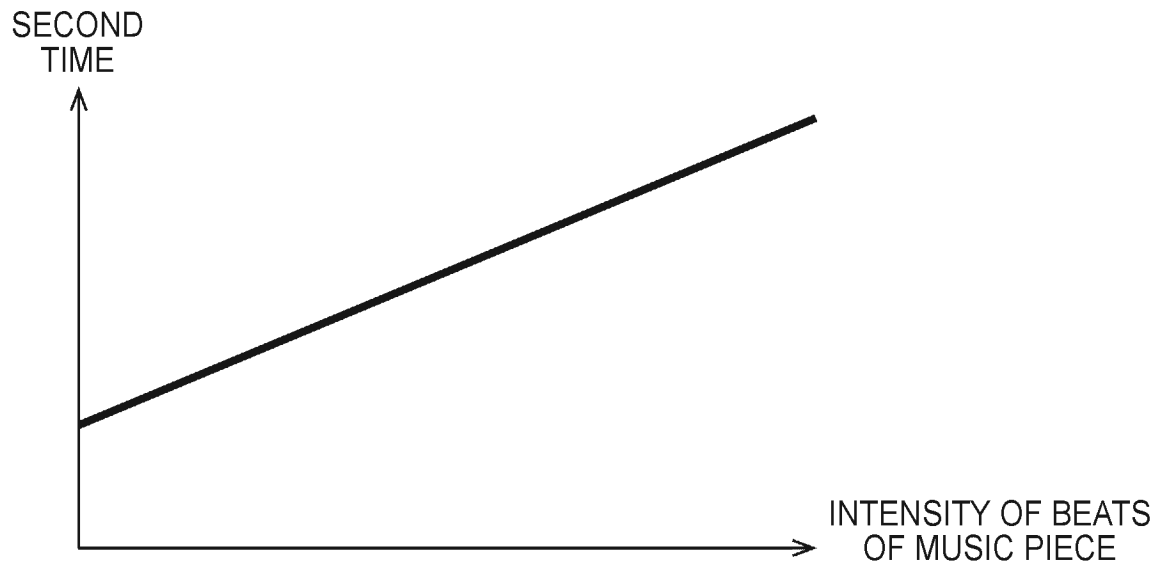


FIG. 9

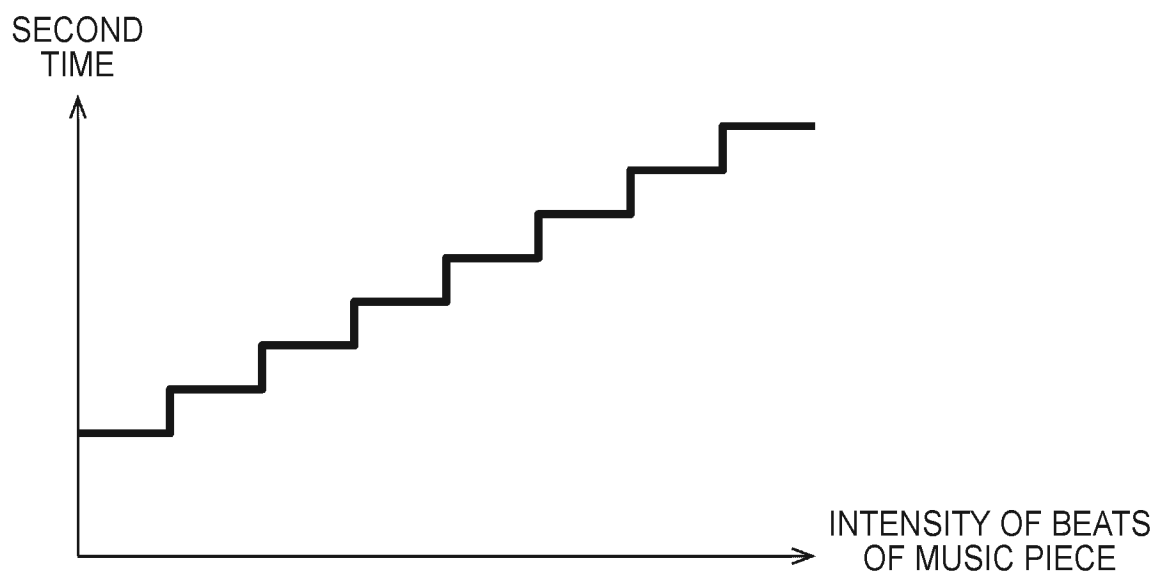
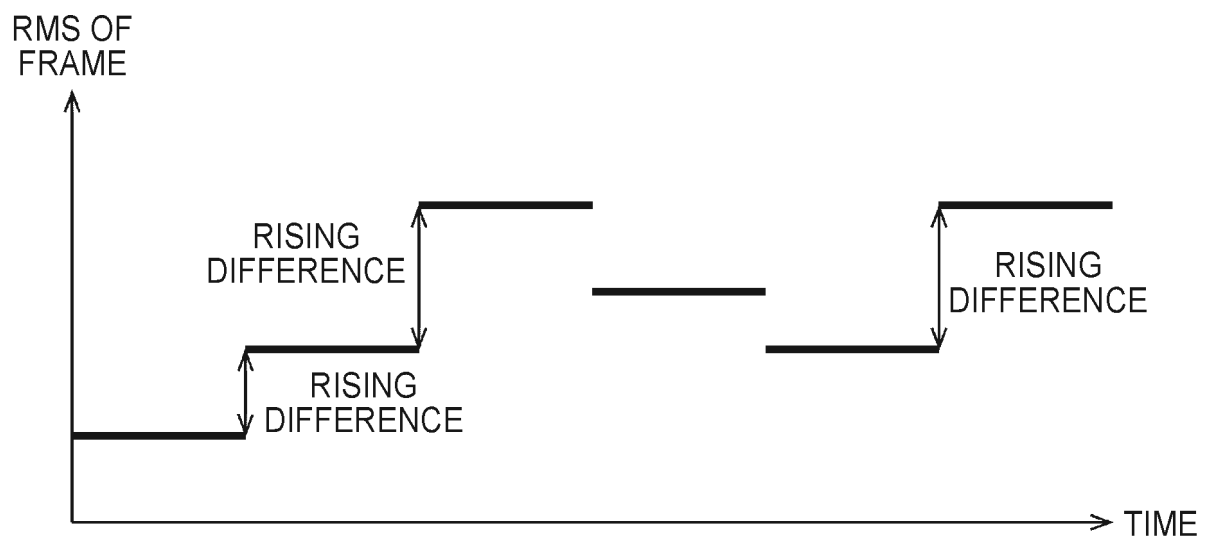


FIG.10



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/043088

A. CLASSIFICATION OF SUBJECT MATTER

G10G 3/04(2006.01)i; *G10L 25/51*(2013.01)i
FI: G10G3/04; G10L25/51 300

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)
G10G3/04; G10L25/51

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

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2024
Registered utility model specifications of Japan 1996-2024
Published registered utility model applications of Japan 1994-2024

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2021-162711 A (PIONEER CORPORATION) 11 October 2021 (2021-10-11) entire text, all drawings	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

“&” document member of the same patent family

Date of the actual completion of the international search

22 January 2024

Date of mailing of the international search report

13 February 2024

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)
3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915
Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2023/043088

5	Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
10	JP 2021-162711 A	11 October 2021	(Family: none)	
15				
20				
25				
30				
35				
40				
45				
50				
55				

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

- JP 2021162707 A [0003]
- WO 2023054236 A [0003]
- WO 2023054237 A [0003]
- JP 2023050570 A [0003]

Non-patent literature cited in the description

- **MASATAKA GOTO ; YOICHI MURAOKA.** A Real-time Beat Tracking System for Audio Signals (Beat Tracking for Music without Drum-sounds). *IEICE Transactions D*, February 1998, vol. J81-D2, 227-237 [0061]