

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2018年11月29日(29.11.2018)



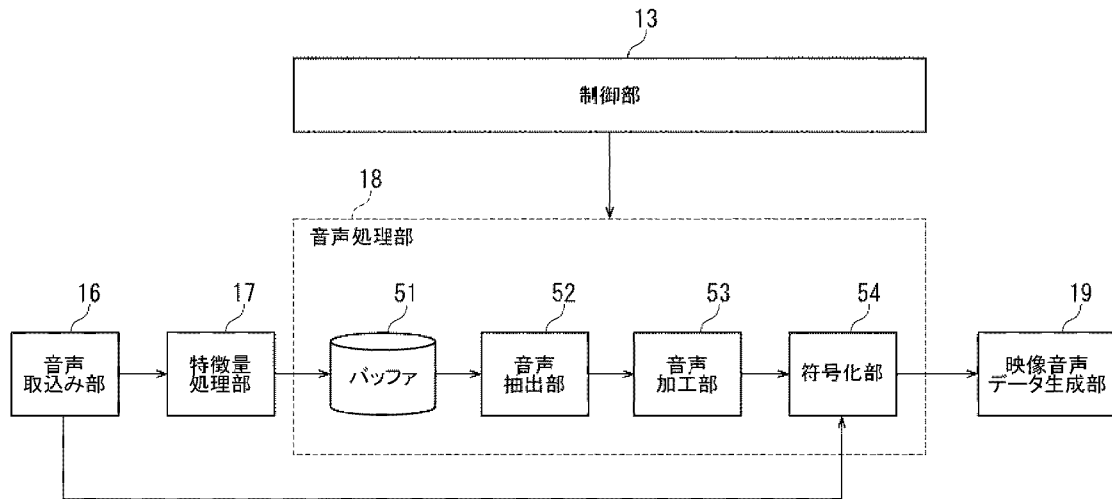
(10) 国際公開番号

WO 2018/216499 A1

- (51) 国際特許分類:
H04N 5/915 (2006.01) H04N 5/77 (2006.01)
G10L 21/055 (2013.01) H04N 5/926 (2006.01)
G11B 20/10 (2006.01) H04N 21/439 (2011.01)
- (21) 国際出願番号: PCT/JP2018/018246
- (22) 国際出願日: 2018年5月11日(11.05.2018)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2017-104132 2017年5月26日(26.05.2017) JP
- (71) 出願人: ソニーセミコンダクタソリューションズ株式会社(SONY SEMICONDUCTOR SOLUTIONS CORPORATION) [JP/JP]; 〒2430014
- 神奈川県厚木市旭町四丁目14番1号 Kanagawa (JP).
- (72) 発明者: 早川 知伸(HAYAKAWA Tomonobu); 〒2430014 神奈川県厚木市旭町四丁目14番1号 ソニーセミコンダクタソリューションズ株式会社内 Kanagawa (JP). 何 信瑩(HO Hsingying); 〒1080075 東京都港区港南1丁目7番1号 ソニー株式会社内 Tokyo (JP).
- (74) 代理人: 西川 孝, 外(NISHIKAWA Takashi et al.); 〒1600023 東京都新宿区西新宿7丁目5番25号 西新宿プライムスクエア9階 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ,

(54) Title: DATA PROCESSING DEVICE, DATA PROCESSING METHOD, PROGRAM, AND DATA PROCESSING SYSTEM

(54) 発明の名称: データ処理装置、データ処理方法、プログラム、及び、データ処理システム



- 13 Control unit
16 Audio capturing unit
17 Feature quantity processing unit
18 Audio processing unit
19 Video/audio data generation unit
51 Buffer
52 Audio extraction part
53 Audio manipulation part
54 Coding part

(57) Abstract: The present technology relates to a data processing device, a data processing method, a program, and data processing system for enabling meaningful audio to be recorded for intermittently-captured images. The data processing device is provided with an audio extraction part for extracting, from audio data corresponding to audio captured over a period during which the multiple images were intermittently captured, one or more audio blocks to be reproduced along with video data drawn from multiple images, on the basis of a prescribed feature quantity. The present technology can be

BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告(条約第21条(3))

applied, for example, to a recording device for interval shooting.

(57) 要約: 本技術は、間欠的な画像に対して有意な音声を記録することができるようにするデータ処理装置、データ処理方法、プログラム、及び、データ処理システムに関する。データ処理装置は、間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する音声データから、前記複数の画像に基づく映像データとともに再生する1以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出部を備える。本技術は、例えば、インターバル撮影を行う記録装置に適用できる。

明 細 書

発明の名称：

データ処理装置、データ処理方法、プログラム、及び、データ処理システム

技術分野

[0001] 本技術は、データ処理装置、データ処理方法、プログラム、及び、データ処理システムに関し、特に、間欠的な画像を記録する場合に用いて好適なデータ処理装置、データ処理方法、プログラム、及び、データ処理システムに関する。

背景技術

[0002] 従来、所定の間隔で間欠的に画像を撮影するインターバル撮影を行うことが可能な装置が普及している。

[0003] また、従来、映像を記録する時間帯と音声を記録する時間帯とを個別に設定し、時間帯毎に映像と音声の両方を記録したり、いずれか一方のみを記録したりすることが可能な映像音声記録装置が提案されている（例えば、特許文献1参照）。

先行技術文献

特許文献

[0004] 特許文献1：特開2007-166042号公報

発明の概要

発明が解決しようとする課題

[0005] しかしながら、インターバル撮影時に画像と音声を同期して取り込んだ場合、音声の間欠的に取り込まれるため、必ずしも有意な音声を記録できるとは限らない。また、1フレーム分の画像の記録時間は非常に短いため、画像と同期して記録した音声の内容を人が認識することは困難である。

[0006] また、特許文献1に記載の発明では、インターバル撮影において有意な音

声を記録できるようにすることは考慮されていない。

[0007] 本技術は、このような状況に鑑みてなされたものであり、間欠的な画像に対して有意な音声を記録できるようにするものである。

課題を解決するための手段

[0008] 本技術の第1の側面のデータ処理装置は、間欠的な複数の画像が取り込まれた第1の期間内に取り込まれた音声に対応する第1の音声データから、前記複数の画像に基づく映像データとともに再生する1以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出部を備える。

[0009] 本技術の第1の側面のデータ処理方法は、間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する音声データから、前記複数の画像に基づく映像データとともに再生する1以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出ステップを含む。

[0010] 本技術の第1の側面のプログラムは、間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する音声データから、前記複数の画像に基づく映像データとともに再生する1以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出ステップを含む処理をコンピュータに実行させる。

[0011] 本技術の第2の側面のデータ処理システムは、間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する第1の音声データから、前記複数の画像に基づく映像データとともに再生する1以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出部と、抽出された前記音声ブロックを加工して、前記映像データと同じ長さの第2の音声データを生成する音声加工部と、前記映像データと前記第2の音声データとを多重化し、映像音声データを生成する映像音声データ生成部とを備え、前記音声抽出部、前記音声加工部、及び、前記映像音声データ生成部が、複数の装置に分散して設けられている。

[0012] 本技術の第1の側面においては、間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する音声データから、前記複数の画像に基づく映像データとともに再生する1以上の音声ブロックが所定の特徴量に基づい

て抽出される。

[0013] 本技術の第2の側面においては、間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する第1の音声データから、前記複数の画像に基づく映像データとともに再生する1以上の音声ブロックが所定の特徴量に基づいて抽出され、抽出された前記音声ブロックが加工され、前記映像データと同じ長さの第2の音声データが生成され、前記映像データと前記第2の音声データとが多重化され、映像音声データが生成される。

発明の効果

[0014] 本技術の第1の側面又は第2の側面によれば、間欠的な画像に対して有意な音声を記録することができる。

[0015] なお、ここに記載された効果は必ずしも限定されるものではなく、本開示中に記載された何れかの効果であってもよい。

図面の簡単な説明

[0016] [図1]本技術を適用した記録装置の一実施の形態を示すブロック図である。
[図2]特徴量処理部の構成例を示すブロック図である。
[図3]音声処理部の構成例を示すブロック図である。
[図4]記録処理を説明するためのフローチャートである。
[図5]記録処理を説明するためのフローチャートである。
[図6]画像と音声を同時に間欠的に取り込む場合を説明するための図である。
[図7]本技術を用いて画像と音声を取り込む場合を説明するための図である。
[図8]本技術を適用した編集装置の一実施の形態を示すブロック図である。
[図9]コンピュータの構成例を示す図である。

発明を実施するための形態

[0017] 以下、本技術を実施するための形態について説明する。説明は以下の順序で行う。

1. 第1の実施の形態（記録装置の例）
2. 第2の実施の形態（編集装置の例）
3. 変形例

4. その他

[0018] <<1. 第1の実施の形態>>

まず、図1乃至図7を参照して、本技術の第1の実施の形態について説明する。

[0019] <記録装置の構成例>

図1は、本技術を適用したデータ処理装置の第1の実施の形態である記録装置10の構成例を示すブロック図である。

[0020] 記録装置10は、画像及び音声を取り込み、取り込んだ画像及び音声に基づく映像音声データを生成し、記録する装置である。

[0021] また、記録装置10は、通常記録モード及びインターバル記録モードの少なくとも2つのモードで映像音声データの記録を行うことができる。通常記録モードは、高いフレームレート（例えば、30fps）で連続的に画像を取り込み、映像音声データを記録するモードである。インターバル記録モードは、低いフレームレート（例えば、1fps）で、所定の間隔を空けて間欠的に画像を取り込み、映像音声データを記録するモードである。

[0022] 記録装置10は、入力部11、センサ部12、制御部13、画像取込み部14、画像処理部15、音声取込み部16、特徴量処理部17、音声処理部18、映像音声データ生成部19、及び、記録部20を備える。

[0023] 入力部11は、記録装置10の操作を行うための操作デバイスや、データを入力するための入力デバイスを備える。入力部11は、ユーザ操作に対応する操作信号や、外部から入力されたデータを制御部13に供給する。

[0024] センサ部12は、記録装置10の処理に必要な各種のセンサを備える。例えば、センサ部12は、加速度センサ、ジャイロセンサ、振動センサ、人感センサ等を備える。センサ部12は、各センサから出力されるセンサデータをデータを取得した時刻情報とともに制御部13に供給する。

[0025] 制御部13は、例えば、CPU等の各種のプロセッサ等を備え、記録装置10の各部の制御を行う。

[0026] 画像取込み部14は、例えば、カメラモジュール等を備え、画像の取込み

を行う。画像取込み部 14 は、取り込んだ画像に対応する画像データを画像処理部 15 に供給する。

[0027] また、画像取込み部 14 は、高いフレームレート（例えば、30 f p s）で連続的に画像を取り込む連続撮影、及び、低いフレームレート（例えば、1 f p s）で間欠的に画像を取り込むインターバル撮影（または、タイムラプス撮影）が可能である。

[0028] 画像処理部 15 は、画像取込み部 14 から供給される画像データに対して、必要に応じて各種の画像処理を施して、映像音声データ生成部 19 に供給する。

[0029] 音声取込み部 16 は、例えば、マイクロホン等を備え、音声の取込みを行う。音声取込み部 16 は、取り込んだ音声に対応する音声データを特徴量処理部 17 又は音声処理部 18 に供給する。

[0030] 特徴量処理部 17 は、音声データの特徴量を抽出し、抽出した特徴量に基づいて、音声データのフィルタリング処理を行う。これにより、音声データの一部が音声処理部 18 に供給され、残りが破棄される。

[0031] 音声処理部 18 は、音声取込み部 16 又は特徴量処理部 17 から供給される音声データに対して、必要に応じて各種の音声処理を施して、映像音声データ生成部 19 に供給する。

[0032] 映像音声データ生成部 19 は、画像処理部 15 から供給される画像データに基づく映像データと、音声処理部 18 から供給される音声データに基づいて、映像音声データを生成し、記録部 20 に記録させる処理を行う。

[0033] 記録部 20 は、例えば、磁気ディスク、光ディスク、光磁気ディスク、又は、半導体メモリ等の記録媒体を備える。

[0034] <特徴量処理部の構成例>

図 2 は、図 1 の特徴量処理部 17 の構成例を示すブロック図である。

[0035] 特徴量処理部 17 は、音圧検出部 31 及び特徴量抽出部 32 を備える。

[0036] 音圧検出部 31 は、音声取込み部 16 から供給される音声データの特徴量の 1 つである音圧を検出する。音圧検出部 31 は、音圧の検出結果に基づい

て、音声データの一部を特徴量抽出部 32 に供給し、残りを破棄する。

[0037] 特徴量抽出部 32 は、音圧検出部 31 から供給される音声データを所定の長さの音声ブロックに分割するとともに、各音声ブロックの音圧以外の 1 つ以上の視点に基づく 1 次元以上の特徴量を抽出する。また、特徴量抽出部 32 は、抽出した特徴量に基づくスコアを各音声ブロックに付与し、そのスコアに基づいて、一部の音声ブロックを特徴量処理部 17 に供給し、残りの音声ブロックを破棄する。

[0038] <音声処理部の構成例>

図 3 は、図 1 の音声処理部 18 の構成例を示すブロック図である。

[0039] 音声処理部 18 は、バッファ 51、音声抽出部 52、音声加工部 53、及び、符号化部 54 を備える。

[0040] バッファ 51 は、特徴量処理部 17 から供給される音声ブロックを一時的に蓄積する。

[0041] 音声抽出部 52 は、バッファ 51 に蓄積されている音声ブロックの中から、各音声ブロックに付与されている特徴量に基づくスコアに基づいて、一部の音声ブロックを抽出し、音声加工部 53 に供給する。

[0042] 音声加工部 53 は、音声抽出部 52 により抽出された音声ブロックに対して各種の加工処理を行うことにより、映像音声データに多重化して記録する音声データを生成する。音声加工部 53 は、生成した音声データを符号化部 54 に供給する。

[0043] 符号化部 54 は、音声取込み部 16 又は音声加工部 53 から供給される音声データに対して所定の符号化処理を行い、符号化した音声データを映像音声データ生成部 19 に供給する。

[0044] <記録処理>

次に、図 4 及び図 5 のフローチャートを参照して、記録装置 10 により実行される記録処理について説明する。なお、ここでは、記録装置 10 が音声を取り込み、記録する処理を中心に説明し、その他の処理は必要に応じて適宜説明する。

- [0045] この処理は、例えば、記録処理の開始の指示が入力部 11 を介して制御部 13 に入力されたとき開始される。或いは、この処理は、例えば、記録処理のスケジュールが予め設定されており、その開始時刻になったとき開始される。また、記録処理の開始時まで、通常記録モード又はインターバル記録モードの選択が行われる。
- [0046] ステップ S1 において、音声取込み部 16 は、音声の取込みを開始する。
- [0047] なお、このとき、画像取込み部 14 は、所定のフレームレートで画像の取込みを行い、取り込んだ画像に対応する画像データを画像処理部 15 に供給する処理を開始する。画像を取り込むフレームレートは、通常記録モードに設定されている場合、例えば、30 fps に設定され、インターバル記録モードに設定されている場合、例えば、1 fps に設定される。
- [0048] また、画像処理部 15 は、符号化等の所定の画像処理を画像データに施して、映像音声データ生成部 19 に供給する処理を開始する。
- [0049] なお、通常記録モード又はインターバル記録モードのいずれの場合においても、音声は連続して取り込まれる。
- [0050] ステップ S2 において、制御部 13 は、通常記録モードに設定されているか否かを判定する。通常記録モードに設定されていると判定された場合、処理はステップ S3 に進む。
- [0051] ステップ S3 において、記録装置 10 は、符号化処理を開始する。具体的には、音声取込み部 16 は、制御部 13 の制御の下に、連続して取り込んだ音声に対応する音声データを符号化部 54 に供給する処理を開始する。符号化部 54 は、音声データに対して所定の符号化処理を行い、符号化後の音声データを映像音声データ生成部 19 に供給する処理を開始する。
- [0052] ステップ S4 において、映像音声データ生成部 19 は、映像音声データの生成及び記録を開始する。例えば、以下の処理が開始される。
- [0053] 映像音声データ生成部 19 は、画像処理部 15 から供給される各フレームの画像データに基づいて、所定のフォーマットのビットストリームである映像データを生成する。また、映像音声データ生成部 19 は、符号化部 54 か

ら供給される音声データを所定のフォーマットのビットストリームに変換する。さらに、映像音声データ生成部 19 は、映像データと音声データを多重化することにより映像音声データを生成し、記録部 20 に記録させる。

[0054] ステップ S5 において、制御部 13 は、インターバル記録モードに設定されたか否かを判定する。インターバル記録モードに設定されていないと判定された場合、処理はステップ S6 に進む。

[0055] ステップ S6 において、制御部 13 は、記録処理を終了するか否かを判定する。記録処理を終了しないと判定された場合、処理はステップ S5 に戻る。

[0056] その後、ステップ S5 において、インターバル記録モードに設定されたと判定されるか、ステップ S6 において、記録処理を終了すると判定されるまで、ステップ S5 及びステップ S6 の判定処理が繰り返し実行され、通常記録モードによる映像音声データの記録が継続する。

[0057] 一方、ステップ S6 において、制御部 13 は、例えば、記録処理の停止の指示が入力部 11 を介して入力された場合、記録処理を終了すると判定する。或いは、制御部 13 は、例えば、記録処理のスケジュールが予め設定されており、その終了時刻になったとき、記録処理を終了すると判定する。そして、記録処理を終了すると判定された場合、記録処理は終了する。

[0058] また、ステップ S5 において、制御部 13 は、例えば、インターバル記録モードの設定の指示が入力部 11 を介して入力された場合、インターバル記録モードに設定されたと判定し、処理はステップ S7 に進む。これは、通常記録モードからインターバル記録モードに変更された場合である。

[0059] さらに、ステップ S2 において、インターバル記録モードに設定されていると判定された場合、処理はステップ S7 に進む。

[0060] ステップ S7 において、記録装置 10 は、音圧の検出を開始する。具体的には、音声取込み部 16 は、制御部 13 の制御の下に、連続して取り込んだ音声を示す音声データを音圧検出部 31 に供給する処理を開始する。音圧検出部 31 は、音声取込み部 16 から供給される音声データの音圧を検出する

処理を開始する。

- [0061] ステップS 8において、音圧検出部3 1は、音圧に基づく音声データのフィルタリングを開始する。例えば、音圧検出部3 1は、音圧が所定の閾値以上となる状態が所定の時間以上継続する区間の音声データを特徴量抽出部3 2に供給し、それ以外の区間の音声データを破棄する処理を開始する。
- [0062] この区間の長さは、例えば、音声データに含まれる短時間のノイズを除去することが可能な時間に設定される。
- [0063] この閾値は、例えば、有意な音声、すなわち記録したい音声が含まれている可能性が低い音声を破棄することが可能なレベルに設定される。何を有意な音声とするかは、例えば、記録するシーンや目的等により変化する。例えば、植物の成長を記録する場合、その植物の周辺の人の声等が有意な音声とされる。例えば、工事の進捗を記録したい場合、工事現場の音等が有意な音声とされる。例えば、海、山、空等の風景の移り変わりを記録したい場合、自然の音や動物や虫の鳴き声等が有意な音声とされる。
- [0064] このフィルタリング処理により、後段の特徴量抽出部3 2の処理を軽減することができる。
- [0065] ステップS 9において、特徴量抽出部3 2は、特徴量の抽出を開始する。例えば、特徴量抽出部3 2は、音圧検出部3 1から供給される音声データを所定の長さの音声ブロックに区切り、各音声ブロックの特徴量を抽出し、抽出した特徴量に基づいて各音声ブロックのスコアリングを行う処理を開始する。
- [0066] ここで、抽出する特徴量の種類は、例えば、記録したい音声の種類に基づいて設定される。例えば、人の声を記録したい場合、人の声らしさを表す特徴量が抽出対象に設定される。また、例えば、特定の人の声のみを記録したい場合、各個人の声を識別可能なレベルの特徴量が抽出対象に設定される。さらに、例えば、特定の内容の発言のみを記録したい場合、文章解析の結果が特徴量と抽出される。
- [0067] また、音声ブロックの長さは、例えば、音声ブロックに含まれる音声が人

にとって意味のある音声として認識可能な最低限の長さ以上に設定される。

[0068] なお、抽出する特徴量は、２種類（２次元）以上であってもよい。

[0069] また、例えば、記録したい音声の種類によっては、ステップＳ７及びステップＳ８の処理を省略し、全ての音声データの特徴量を抽出するようにしてもよい。

[0070] ステップＳ１０において、特徴量抽出部３２は、特徴量に基づく音声データのフィルタリングを開始する。例えば、特徴量抽出部３２は、特徴量に基づくスコアが所定の閾値以上の音声ブロックをスコアとともにバッファ５１に蓄積し、それ以外の音声ブロックをバッファ５１に蓄積せずに破棄する処理を開始する。

[0071] このフィルタリング処理により、バッファ５１の容量を削減することができる。

[0072] ステップＳ１１において、制御部１３は、映像音声データを記録するタイミングであるか否かを判定する。映像音声データを記録するタイミングでないと判定された場合、処理はステップＳ１２に進む。

[0073] ステップＳ１２において、制御部１３は、通常記録モードに設定されたか否かを判定する。通常記録モードに設定されていないと判定された場合、処理はステップＳ１３に進む。

[0074] ステップＳ１３において、ステップＳ６の処理と同様に、記録処理を終了するか否かが判定され、記録処理を終了しないと判定された場合、処理はステップＳ１１に戻る。

[0075] その後、ステップＳ１１において、映像音声データを記録するタイミングであると判定されるか、ステップＳ１２において、通常記録モードに設定されたと判定されるか、ステップＳ１３において、記録処理を終了すると判定されるまで、ステップＳ１１乃至ステップＳ１３の処理が繰り返し実行される。

[0076] 一方、ステップＳ１３において、記録処理を終了すると判定された場合、処理はステップＳ１４に進む。

- [0077] また、ステップS 1 2において、例えば、制御部 1 3は、通常記録モードの設定の指示が入力部 1 1を介して入力された場合、通常記録モードに設定されたと判定し、処理はステップS 1 4に進む。これは、インターバル記録モードから通常記録モードに変更された場合である。
- [0078] さらに、ステップS 1 1において、例えば、制御部 1 3は、インターバル記録モードによる記録処理が開始されてから、まだ映像音声データの記録が 1 回も行われていない場合、インターバル記録モードによる記録処理の開始から所定の時間が経過したとき、映像音声データを記録するタイミングであると判定する。また、例えば、制御部 1 3は、インターバル記録モードによる記録処理が開始されてから、1 回以上映像音声データの記録が行われている場合、前回映像音声データが記録されてから所定の時間が経過したとき、映像音声データを記録するタイミングであると判定する。そして、映像音声データを記録するタイミングであると判定された場合、処理はステップS 1 1に進む。
- [0079] これにより、インターバル記録モードによる記録処理が継続している期間中に、所定の間隔で定期的に映像音声データが生成され、記録されるようになる。この間隔は、例えば、ユーザが設定するようにしてもよいし、或いは、記録装置 1 0が自動的に設定するようにしてもよい。
- [0080] また、映像音声データを記録する間隔が短くなるほど、映像データと音声データの時間のズレを小さくし、映像データと音声データの相関度を高めることができる。また、バッファ 5 1の容量を小さくし、音声処理部 1 8の処理を分散することができる。一方、映像音声データを記録する間隔を短くしすぎると、例えば、音声短すぎたり、途中で途切れたりして、有意な音声記録できなくなる場合がある。
- [0081] 一方、ステップS 1 1の判定処理を省略することも可能である。この判定処理を省略した場合、インターバル記録モードによる記録処理の終了時に、まとめて映像音声データが生成され、記録されるようになる。
- [0082] ステップS 1 4において、制御部 1 3は、映像記録時間を計算する。具体

的には、制御部 13 は、インターバル記録モードによる記録処理が開始されてから、まだ映像音声データの記録が 1 回も行われていない場合、インターバル記録モードによる記録処理が開始されてから現在までの期間（以下、記録対象期間と称する）の長さを計算する。一方、制御部 13 は、インターバル記録モードによる記録処理が開始されてから、1 回以上映像音声データの記録が行われている場合、前回映像音声データが記録された時点から現在までの期間（記録対象期間）の長さを計算する。

[0083] そして、制御部 13 は、記録対象期間内に取り込まれた画像データの長さを映像記録時間として計算する。例えば、記録対象期間の長さが 1 時間であり、画像の取込みが 1 f p s のフレームレートで行われ、1 フレームの画像データの長さが 33 m s である場合、映像記録時間は 118.8 秒（ $= 33 \text{ m s} \times 3600 \text{ フレーム}$ ）となる。

[0084] ステップ S 15 において、音声抽出部 52 は、特徴量に基づいて、音声ブロックを抽出する。例えば、音声抽出部 52 は、バッファ 51 に蓄積されている音声ブロックのうち時間的に連続する音声ブロックを 1 つの音声ブロックにまとめる。また、音声抽出部 52 は、例えば、1 つにまとめた音声ブロックのスコアとして、各音声ブロックのスコアの平均値を計算する。

[0085] そして、音声抽出部 52 は、上記の 1 つにまとめる処理を行った後、バッファ 51 に蓄積されている音声ブロックの中から、スコアが所定の閾値以上の音声ブロックを抽出する。音声抽出部 52 は、抽出した音声ブロックをスコアとともに音声加工部 53 に供給する。また、音声抽出部 52 は、バッファ 51 に蓄積されている音声ブロックを削除し、バッファ 51 をリセットする。

[0086] ステップ S 16 において、音声加工部 53 は、抽出された音声ブロックに基づいて、映像記録時間と同じ長さの音声データを生成する。

[0087] 例えば、音声加工部 53 は、抽出された音声ブロックを取り込まれた時間順に並べて接続する。このとき、音声加工部 53 は、例えば、ユーザが聞いて違和感が生じないように、各音声ブロックにフェード処理（フェードイン

及びフェードアウト)を施す。

[0088] また、音声加工部53は、抽出された音声ブロックの合計時間が映像記録時間より長い場合、接続後の音声データが映像記録時間と同じ長さになるように、音声データの時間方向の圧縮を行う。例えば、音声加工部53は、話速変換により各音声ブロックの再生速度を上げる。なお、再生速度の最大値は、人が会話を聞き取れるレベル（例えば、2倍速）に制限される。また、例えば、音声加工部53は、クロスフェード処理により、隣接する音声ブロックを違和感なく重畳させる。

[0089] なお、音声加工部53は、以上の処理を施したとしても、音声データの長さが映像記録時間より長い場合、例えば、スコアが低いものから順に音声ブロックを破棄し、音声データの長さが映像記録時間と等しくなるように調整する。

[0090] 一方、音声加工部53は、抽出された音声ブロックの合計時間が映像記録時間より短い場合、1以上の無音の音声ブロック（以下、無音ブロックと称する）を挿入することにより、音声データの長さを映像記録時間に合わせる。このとき、例えば、各音声ブロックと、各音声ブロックが取り込まれた時刻付近に取り込まれた画像データとの時間のズレが小さくなるように、無音ブロックの挿入位置が設定される。

[0091] また、音声加工部53は、必要に応じて、音声データのゲイン調整やノイズ除去等の処理を行う。

[0092] そして、音声加工部53は、生成した音声データを符号化部54に供給する。

[0093] ステップS17において、符号化部54は、符号化処理を行う。具体的には、符号化部54は、音声加工部53から供給された音声データに対して所定の符号化処理を行い、符号化後の音声データを映像音声データ生成部19に供給する。

[0094] ステップS18において、映像音声データ生成部19は、映像音声データを生成し、記録する。例えば、映像音声データ生成部19は、記録対象期間

内に取り込まれた各フレームの画像データ（間欠的な画像データ）に基づいて、所定のフォーマットのビットストリームである映像データを生成する。また、映像音声データ生成部 19 は、符号化部 54 から供給された音声データを所定のフォーマットのビットストリームに変換する。このとき、通常記録モードで記録された映像データ及び音声データと同じように間欠的に取り込まれた画像を連続して再生するように、映像データと音声データの時間情報（例えば、タイムスタンプ等）の変更が行われる。さらに、映像音声データ生成部 19 は、映像データと音声データを多重化することにより映像音声データを生成し、記録部 20 に記録させる。

[0095] ステップ S 19 において、ステップ S 6 の処理と同様に、記録処理を終了するか否かが判定され、記録処理を終了しないと判定された場合、処理はステップ S 20 に進む。

[0096] ステップ S 20 において、制御部 13 は、インターバル記録モードに設定されているか否かを判定する。インターバル記録モードに設定されていると判定された場合、処理はステップ S 11 に戻り、ステップ S 11 以降の処理が実行される。これは、インターバル記録モードによる記録処理中に、所定のタイミングで映像音声データの生成及び記録が行われた場合であり、引き続き、インターバル記録モードによる記録処理が継続される。

[0097] 一方、ステップ S 20 において、通常記録モードに設定されていると判定された場合、処理はステップ S 3 に戻り、ステップ S 3 以降の処理が実行される。これは、インターバル記録モードから通常記録モードに変更された場合であり、その後、通常記録モードによる記録処理が開始される。

[0098] また、ステップ S 19 において、記録処理を終了すると判定された場合、記録処理は終了する。

[0099] 以上のようにして、間欠的な画像に対して有意な音声を記録できるようになる。

[0100] 例えば、図 6 は、花の成長を間欠的に記録する場合に、画像と音声を同時に取り込むようにしたときの画像と音声を取り込むタイミングの例を示して

いる。図内の1行目は、花が成長する様子を示しており、2行目は、画像を取り込むタイミングを、カメラの絵を含む枠で示し、3行目は、音声を取り込むタイミングを、波形図を含む枠で示している。

[0101] 例えば、33msの長さの画像が所定の間隔で取り込まれ、記録される。また、画像と同期して、33msの長さの音声も所定の間隔で取り込まれ、記録される。しかし、画像を取り込んでいるときに、必ずしも有意な音声も背景に流れているとは限らない。また、33msの長さの音声は、人にとって有意な音声として認識することが困難である。

[0102] 一方、図7は、花の成長を間欠的に記録する場合に、本技術を用いたときの画像と音声を取り込むタイミングの例を、図6と同様に示している。なお、図7では、音声を取り込むタイミングを示した行の下に、取り込まれる音声の具体例が示されている。

[0103] 本技術を用いた場合、画像の取込みとは同期せずに、例えば花を観察している子供の声等の有意な音声のみが抽出され記録されるため、映像音声データの価値が向上する。そして、間欠的に取り込まれた複数のフレームの画像からなる映像と合わせて、抽出された音声も再生されるため、臨場感が増す。

[0104] また、有意な音声のみ記録され、不要な音声は破棄されるため、映像音声データの容量が増大することが防止される。

[0105] さらに、再生環境や装置に依存することなく、インターバル記録モードで記録された映像音声データを、通常記録モードで記録された映像音声データと同様の方法により再生することができる。

[0106] また、上述したように、通常記録モードとインターバル記録モードの切り替えをシームレスに行うことができる。また、途中で記録モードが切り替わっても、再生時に特別な処理を行うことなく連続して再生することが可能である。

[0107] さらに、抽出する特徴量の種類を変更することにより、記録するシーンや目的に適した有意な音声を適切に抽出することが可能になる。

[0108] なお、間欠的に取り込まれた画像を連続して再生する場合、実質的に映像が早送り再生されるようになるため、映像と音声の間に少々の時間のズレが生じて、ユーザはあまり気にならない。また、話速変換により音声の再生速度を上げて、ユーザは、あまり違和感を覚えない。

[0109] << 2. 第2の実施の形態 >>

次に、図8を参照して、本技術の第2の実施の形態について説明する。

[0110] <編集装置の構成例>

図8は、本技術を適用したデータ処理装置の第2の実施の形態である編集装置100の構成例を示すブロック図である。なお、図中、図1の記録装置10と対応する部分には同じ符号を付してあり、その説明は適宜省略する。

[0111] 編集装置100は、例えば、画像が通常のフレームレートで連続的に記録された映像音声データから、画像が間欠的に記録された映像音声データを生成する装置である。

[0112] 編集装置100は、記録装置10と比較して、制御部13の代わりに制御部101が設けられ、分離部102及び画像抽出部103が追加され、センサ部12、画像取込み部14、及び、音声取込み部16が削除されている点異なる。

[0113] 制御部101は、例えば、CPU等の各種のプロセッサ等を備え、編集装置100の各部の制御を行う。

[0114] 分離部102は、外部から供給される映像音声データに多重化されている映像データと音声データを分離する。そして、分離部102は、映像データを画像抽出部103に供給し、音声データを特徴量処理部17に供給する。

[0115] 画像抽出部103は、映像データに含まれる複数のフレームの映像データを所定の間隔で抽出し、抽出した映像データを画像処理部15に供給する。これにより、上述した記録装置10がインターバル撮影モードにより間欠的に取り込んだ映像データと同様の間欠的な映像データが、画像処理部15に供給されるようになる。そして、画像処理部15は、各映像データに対して所定の画像処理を行い、映像音声データ生成部19に供給する。

[0116] 特徴量抽出部 17 及び音声処理部 18 は、分離部 102 から供給される音声データに基づいて、上述した記録装置 10 のインターバル撮影モード時の処理と同様の処理により、音声データを生成し、映像音声データ生成部 19 に供給する。

[0117] 映像音声データ生成部 19 は、上述した記録装置 10 のインターバル撮影モード時の処理と同様の処理により、間欠的な画像データに基づいて映像データを生成する。そして、映像音声データ生成部 19 は、映像データと音声データに基づいて映像音声データを生成し、記録部 20 に記録させる。

[0118] このようにして、画像が連続的に記録された映像音声データから、画像が間欠的に記録された映像音声データが生成される。また、すでに記録されている音声データの中から有意な音声を含む音声データが抽出され、映像音声データに記録される。

[0119] <<3. 変形例>>

以下、上述した本開示に係る技術の実施の形態の変形例について説明する。

[0120] <システムの構成に関する変形例>

以上の説明では、図 1 の記録装置 10 及び図 8 の編集装置 100 が、それぞれ 1 台で全ての処理を行う例を示したが、複数の装置（或いは、複数の LSI、プラットフォーム等）で分散して処理を行うようにしてもよい。複数の装置で分散して処理を行う場合、各装置が直接接続されるようにしてもよいし、ネットワーク等を介して間接的に接続されるようにしてもよい。或いは、各装置を直接又は間接的に接続せずに、装置間のデータの授受を、記録媒体等を介して行うようにしてもよい。

[0121] 例えば、画像取込み部 14 及び音声取込み部 16 を記録装置 10 の外部に設けるようにしてもよい。例えば、画像取込み部 14 及び音声取込み部 16 を備える撮影機器を撮影したい場所に設置し、常にインターバル撮影モードで撮影を行い、取り込まれた画像データ及び音声データを記録装置 10 に伝送するようにしてもよい。これにより、撮影機器の負荷を軽減し、消費電力

を削減することができる。また、撮影機器と記録装置 10 の間のデータの伝送量が小さくなり、両者の間の伝送路の伝送容量を小さくすることができる。

[0122] また、例えば、映像音声データ生成部 19 及び記録部 20 を記録装置 10 の外部に設けるようにしてもよい。この場合、記録装置 10 から映像音声データ生成部 19 及び記録部 20 を備える後段の装置への画像データ及び音声データの伝送を間欠的に又は低ビットレートで行うことができ、両者の間の伝送路の伝送容量をさらに削減することができる。

[0123] また、例えば、記録部 20 を記録装置 10 又は編集装置 100 の外部に設けるようにしてもよい。例えば、記録装置 10 又は編集装置 100 は、生成した映像音声データを外部の機器（例えば、ローカルサーバ、スマートフォン等）に供給し、記録させたり、ネットワークを介してクラウドサービス等を提供するサーバに供給し、記録させたりするようにしてもよい。

[0124] さらに、例えば、画像データ、及び、特徴量に基づくスコアを付与した音声データ（以下、スコア付き音声データと称する）を記録装置 10 から外部の機器やサーバに送信し、外部で記録用の音声データを生成し、その音声データを含む映像音声データを生成するようにしてもよい。

[0125] また、例えば、画像データ及びスコア付き音声データを複数の装置に送信し、各装置で異なる音声データを生成し、異なる音声データを含む映像音声データを生成するようにしてもよい。例えば、複数のスマートフォンに画像データ及びスコア付き音声データを送信し、各スマートフォンにおいて、スマートフォンの持ち主の声に基づいて音声データを抽出し、記録するようにしてもよい。

[0126] <処理に関する変形例>

例えば、インターバル記録モードにより映像音声データを記録する期間、又は、インターバル記録モードにおいて映像音声データを生成し、記録する間隔（以下、記録間隔と称する）が事前に決まっている場合、図 5 のステップ S 14 の処理で計算した映像記録時間が事前に定まる。そして、話速変換

及びクロスフェード処理等を考慮して、その映像記録時間の映像データに多重化可能な音声データの時間方向の圧縮前の時間の最大値（以下、音声記録可能時間と称する）が事前に定まる。この場合、例えば、バッファ５１に蓄積されている音声ブロックの合計時間が音声記録可能時間を超えそうな場合、音声ブロックの合計時間が音声記録可能時間を超えないように、スコアが小さい音声ブロックから順に消去していくことにより、必要なバッファ５１の容量を削減することができる。

[0127] また、例えば、まだ記録していない映像データの時間、バッファ５１に蓄積されている音声ブロックの合計時間、及び、各音声ブロックのスコア等に基づいて、確実に抽出して記録されると予測されるスコアの高い音声ブロックに対して、加工処理や符号化処理を前倒しして行うようにしてもよい。これにより、バッファ５１の容量を削減することができる。

[0128] さらに、例えば、スコアが高い音声ブロックを、ほぼリアルタイムに符号化し、所定のフォーマットに変換し、記録部２０に記録させるようにしてもよい。これにより、バッファ５１の容量を削減することができる。なお、最終的な音声データの時間が映像記録時間より長い場合、例えば、音声データの不連続点周辺の音声をデコードした後、クロスフェードしながら再エンコードすることにより、音声データの長さを調整することが可能である。一方、最終的な音声データの時間が映像記録時間より短い場合、例えば、無音ブロックを挿入することにより、音声データの長さを調整することが可能である。

[0129] また、例えば、インターバル記録モードにおいて記録間隔が予め定められている場合、特徴量処理部１７と音声処理部１８の処理を並列に行うことができ、処理時間を短縮することができる。

[0130] さらに、以上の説明では、音声データから抽出された特徴量に基づいて、記録する音声データを抽出する例を示したが、音声データから抽出された特徴量以外の特徴量に基づいて、記録する音声データを抽出するようにしてもよい。

[0131] 例えば、センサ部 12 が備える各種のセンサが音声の取り込みと並行して検出した時系列のセンサデータに基づく特徴量に基づいて、記録する音声データを抽出するようにしてもよい。例えば、人感センサにより周囲に人の存在が検出された期間の音声データを抽出するようにしてもよい。また、例えば、振動センサより所定の閾値以上の振動が検出された期間の音声データを抽出するようにしてもよい。

[0132] なお、音声データから抽出された特徴量とそれ以外の特徴量の両方を用いるようにしてもよい。

[0133] また、例えば、特徴量抽出部 32 が、センサ部 12 から供給されるセンサデータから特徴量を抽出するようにしてもよい。

[0134] また、例えば、図 1 の画像取込み部 14 が、インターバル記録モード時に通常記録モードと同じフレームレートで画像を取り込み、所定の間隔で画像データを抽出して画像処理部 15 に供給するようにしてもよい。

[0135] <本技術の適用例>

本技術は、間欠的な画像に基づく映像データとともに再生する音声データを生成する機能を含む装置（例えば、撮影装置、記録装置、編集装置等）やシステムに適用することができる。

[0136] また、本技術の対象となる映像データは、間欠的に取り込んだ画像に基づく映像データでも、連続的に取り込んだ画像から抽出された画像に基づく映像データでもよい。また、間欠的な画像間の時間間隔は、固定でも可変でもよい。例えば、被写体の動きに合わせて断続的に取り込んだ画像に基づく映像データも、本技術の対象となる。

[0137] <<4. その他>>

<コンピュータの構成例>

上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、コンピュータにインストールされる。ここで、コンピュータには、専用のハードウェア

に組み込まれているコンピュータや、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどが含まれる。

[0138] 図9は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すブロック図である。

[0139] コンピュータにおいて、CPU (Central Processing Unit) 401, ROM (Read Only Memory) 402, RAM (Random Access Memory) 403は、バス404により相互に接続されている。

[0140] バス404には、さらに、入出力インターフェース405が接続されている。入出力インターフェース405には、入力部406、出力部407、記録部408、通信部409、及びドライブ410が接続されている。

[0141] 入力部406は、入力スイッチ、ボタン、マイクロホン、撮像素子などよりなる。出力部407は、ディスプレイ、スピーカなどよりなる。記録部408は、ハードディスクや不揮発性のメモリなどよりなる。通信部409は、ネットワークインターフェースなどよりなる。ドライブ410は、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリなどのリムーバブル記録媒体411を駆動する。

[0142] 以上のように構成されるコンピュータでは、CPU401が、例えば、記録部408に記録されているプログラムを、入出力インターフェース405及びバス404を介して、RAM403にロードして実行することにより、上述した一連の処理が行われる。

[0143] コンピュータ (CPU401) が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブル記録媒体411に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の伝送媒体を介して提供することができる。

[0144] コンピュータでは、プログラムは、リムーバブル記録媒体411をドライブ410に装着することにより、入出力インターフェース405を介して、

記録部408にインストールすることができる。また、プログラムは、有線または無線の伝送媒体を介して、通信部409で受信し、記録部408にインストールすることができる。その他、プログラムは、ROM402や記録部408に、あらかじめインストールしておくことができる。

[0145] なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

[0146] また、本明細書において、システムとは、複数の構成要素（装置、モジュール（部品）等）の集合を意味し、すべての構成要素が同一筐体中にあるか否かは問わない。したがって、別個の筐体に収納され、ネットワークを介して接続されている複数の装置、及び、1つの筐体の中に複数のモジュールが収納されている1つの装置は、いずれも、システムである。

[0147] さらに、本技術の実施の形態は、上述した実施の形態に限定されるものではなく、本技術の要旨を逸脱しない範囲において種々の変更が可能である。

[0148] 例えば、本技術は、1つの機能をネットワークを介して複数の装置で分担、共同して処理するクラウドコンピューティングの構成をとることができる。

[0149] また、上述のフローチャートで説明した各ステップは、1つの装置で実行する他、複数の装置で分担して実行することができる。

[0150] さらに、1つのステップに複数の処理が含まれる場合には、その1つのステップに含まれる複数の処理は、1つの装置で実行する他、複数の装置で分担して実行することができる。

[0151] ＜構成の組み合わせ例＞

本技術は、以下のような構成をとることもできる。

[0152] (1)

間欠的な複数の画像が取り込まれた第1の期間内に取り込まれた音声に対応する第1の音声データから、前記複数の画像に基づく映像データとともに

再生する 1 以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出部を

備えるデータ処理装置。

(2)

抽出された前記音声ブロックを加工して、前記映像データと同じ長さの第 2 の音声データを生成する音声加工部を

さらに備える前記 (1) に記載のデータ処理装置。

(3)

前記音声加工部は、複数の前記音声ブロックが抽出された場合、複数の前記音声ブロックを時間順に並べて接続する

前記 (2) に記載のデータ処理装置。

(4)

前記音声加工部は、各前記音声ブロックのフェード処理を行う

前記 (3) に記載のデータ処理装置。

(5)

前記音声加工部は、抽出された前記音声ブロックの合計時間が前記映像データの時間より長い場合、隣接する前記音声ブロック間でクロスフェード処理を行う

前記 (4) に記載のデータ処理装置。

(6)

前記音声加工部は、抽出された前記音声ブロックの合計時間が前記映像データの時間より長い場合、前記音声ブロックの話速変換を行う

前記 (2) 乃至 (5) のいずれかに記載のデータ処理装置。

(7)

前記音声加工部は、抽出された前記音声ブロックの合計時間が前記映像データの時間より短い場合、無音の音声ブロックを挿入する

前記 (2) 乃至 (6) のいずれかに記載のデータ処理装置。

(8)

前記映像データと前記第 2 の音声データとを多重化し、映像音声データを生成する映像音声データ生成部を

さらに備える前記（2）乃至（7）のいずれかに記載のデータ処理装置。

（9）

前記映像音声データ生成部は、前記複数の画像を連続して再生するように前記映像データ及び前記音声データの時間情報を変更する

前記（8）に記載のデータ処理装置。

（10）

前記特徴量を抽出する特徴量抽出部を

さらに備える前記（1）乃至（9）のいずれかに記載のデータ処理装置。

（11）

前記特徴量は、前記第 1 の音声データの特徴量、及び、前記第 1 の期間内に所定のセンサにより検出されたデータに基づく特徴量のうち少なくとも 1 つを含む

前記（1）乃至（10）のいずれかに記載のデータ処理装置。

（12）

前記音声抽出部は、所定の長さの第 2 の期間毎に、前記第 2 の期間内に取り込まれた音声に対応する音声データから、前記第 2 の期間内に取り込まれた間欠的な複数の画像に基づく映像データとともに再生する 1 以上の前記音声ブロックを前記特徴量に基づいて抽出する

前記（1）乃至（11）のいずれかに記載のデータ処理装置。

（13）

画像の取込みを行う画像取込み部と、

音声の取込みを行う音声取込み部と

をさらに備える前記（1）乃至（12）のいずれかに記載のデータ処理装置。

（14）

間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する

音声データから、前記複数の画像に基づく映像データとともに再生する１以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出ステップを含むデータ処理方法。

(１５)

間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する音声データから、前記複数の画像に基づく映像データとともに再生する１以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出ステップを含む処理をコンピュータに実行させるためのプログラム。

(１６)

間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対応する第１の音声データから、前記複数の画像に基づく映像データとともに再生する１以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出部と、

抽出された前記音声ブロックを加工して、前記映像データと同じ長さの第２の音声データを生成する音声加工部と、

前記映像データと前記第２の音声データとを多重化し、映像音声データを生成する映像音声データ生成部と

を備え、

前記音声抽出部、前記音声加工部、及び、前記映像音声データ生成部が、複数の装置に分散して設けられている

データ処理システム。

(１７)

画像の取込みを行う画像取込み部と、

音声の取込みを行う音声取込み部と

をさらに備える前記（１６）に記載のデータ処理システム。

符号の説明

[0153] １０ 記録装置, １２ センサ部, １３ 制御部, １４ 画像取込み部, １５ 画像処理部, １６ 音声取込み部, １７ 特徴量処理部, １８ 音声処理部, １９ 映像音声データ生成部, ３１ 音圧検出

部, 32 特徴量抽出部, 51 バッファ, 52 音声抽出部, 5
3 音声加工部, 54 符号化部, 100 編集装置, 101 制御
部, 102 分離部, 103 画像抽出部

請求の範囲

- [請求項1] 間欠的な複数の画像が取り込まれた第1の期間内に取り込まれた音声に対応する第1の音声データから、前記複数の画像に基づく映像データとともに再生する1以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出部を
備えるデータ処理装置。
- [請求項2] 抽出された前記音声ブロックを加工して、前記映像データと同じ長さの第2の音声データを生成する音声加工部を
さらに備える請求項1に記載のデータ処理装置。
- [請求項3] 前記音声加工部は、複数の前記音声ブロックが抽出された場合、複数の前記音声ブロックを時間順に並べて接続する
請求項2に記載のデータ処理装置。
- [請求項4] 前記音声加工部は、各前記音声ブロックのフェード処理を行う
請求項3に記載のデータ処理装置。
- [請求項5] 前記音声加工部は、抽出された前記音声ブロックの合計時間が前記映像データの時間より長い場合、隣接する前記音声ブロック間でクロスフェード処理を行う
請求項4に記載のデータ処理装置。
- [請求項6] 前記音声加工部は、抽出された前記音声ブロックの合計時間が前記映像データの時間より長い場合、前記音声ブロックの話速変換を行う
請求項2に記載のデータ処理装置。
- [請求項7] 前記音声加工部は、抽出された前記音声ブロックの合計時間が前記映像データの時間より短い場合、無音の音声ブロックを挿入する
請求項2に記載のデータ処理装置。
- [請求項8] 前記映像データと前記第2の音声データとを多重化し、映像音声データを生成する映像音声データ生成部を
さらに備える請求項2に記載のデータ処理装置。
- [請求項9] 前記映像音声データ生成部は、前記複数の画像を連続して再生する

ように前記映像データ及び前記音声データの時間情報を変更する
請求項 8 に記載のデータ処理装置。

[請求項10] 前記特徴量を抽出する特徴量抽出部を
さらに備える請求項 1 に記載のデータ処理装置。

[請求項11] 前記特徴量は、前記第 1 の音声データの特徴量、及び、前記第 1 の
期間内に所定のセンサにより検出されたデータに基づく特徴量のうち
少なくとも 1 つを含む
請求項 1 に記載のデータ処理装置。

[請求項12] 前記音声抽出部は、所定の長さの第 2 の期間毎に、前記第 2 の期間
内に取り込まれた音声に対応する音声データから、前記第 2 の期間内
に取り込まれた間欠的な複数の画像に基づく映像データとともに再生
する 1 以上の前記音声ブロックを前記特徴量に基づいて抽出する
請求項 1 に記載のデータ処理装置。

[請求項13] 画像の取込みを行う画像取込み部と、
音声の取込みを行う音声取込み部と
をさらに備える請求項 1 に記載のデータ処理装置。

[請求項14] 間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対
応する音声データから、前記複数の画像に基づく映像データとともに
再生する 1 以上の音声ブロックを所定の特徴量に基づいて抽出する音
声抽出ステップを
含むデータ処理方法。

[請求項15] 間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対
応する音声データから、前記複数の画像に基づく映像データとともに
再生する 1 以上の音声ブロックを所定の特徴量に基づいて抽出する音
声抽出ステップを
含む処理をコンピュータに実行させるためのプログラム。

[請求項16] 間欠的な複数の画像が取り込まれた期間内に取り込まれた音声に対
応する第 1 の音声データから、前記複数の画像に基づく映像データと

ともに再生する 1 以上の音声ブロックを所定の特徴量に基づいて抽出する音声抽出部と、

抽出された前記音声ブロックを加工して、前記映像データと同じ長さの第 2 の音声データを生成する音声加工部と、

前記映像データと前記第 2 の音声データとを多重化し、映像音声データを生成する映像音声データ生成部と

を備え、

前記音声抽出部、前記音声加工部、及び、前記映像音声データ生成部が、複数の装置に分散して設けられている

データ処理システム。

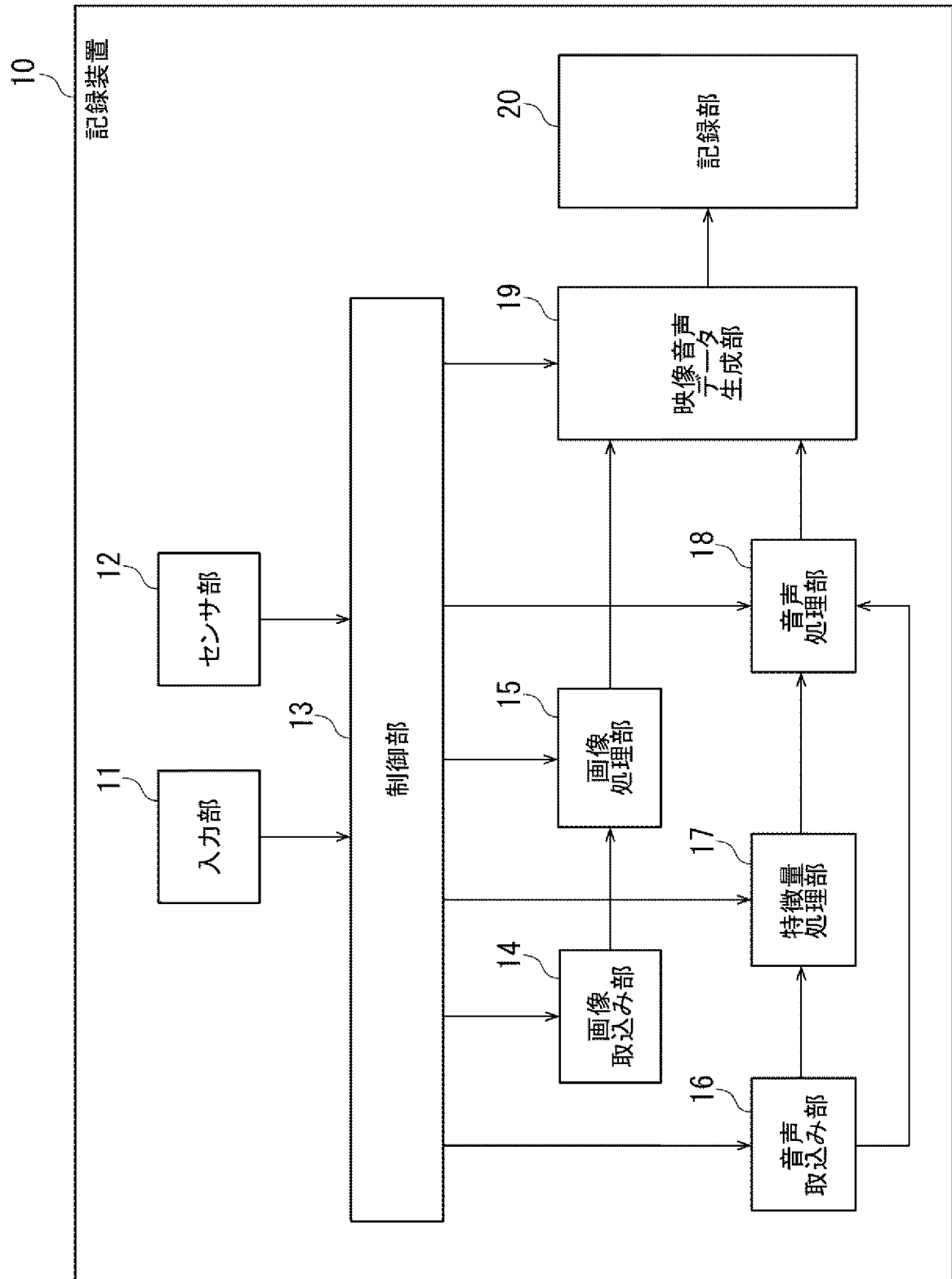
[請求項 17]

画像の取込みを行う画像取込み部と、

音声の取込みを行う音声取込み部と

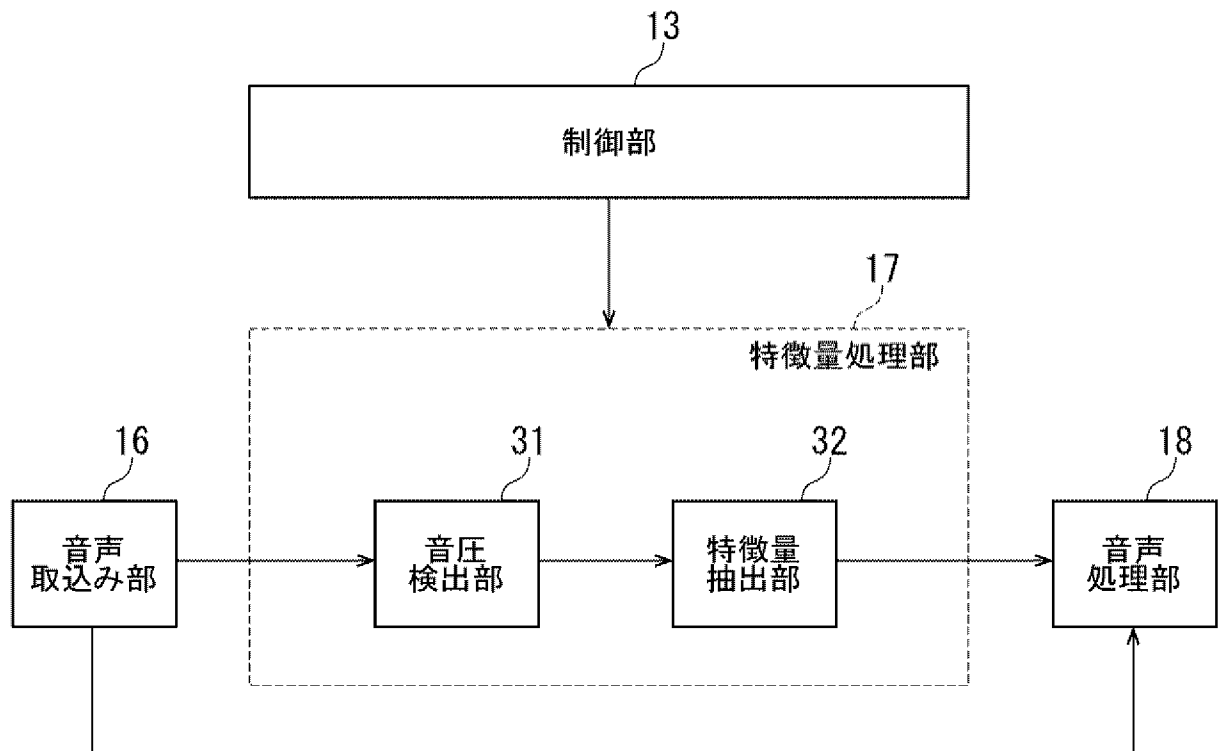
をさらに備える請求項 16 に記載のデータ処理システム。

[図1]
FIG. 1

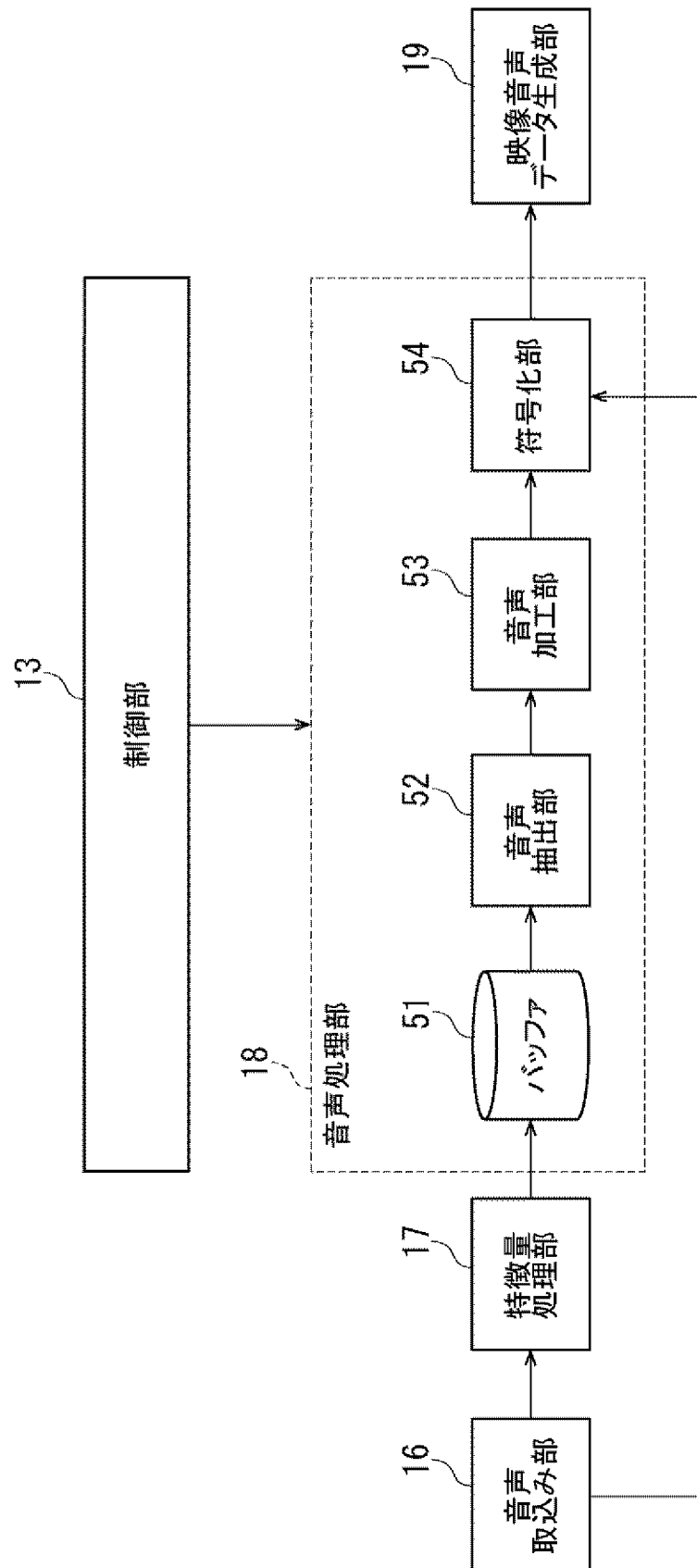


[図2]

FIG. 2

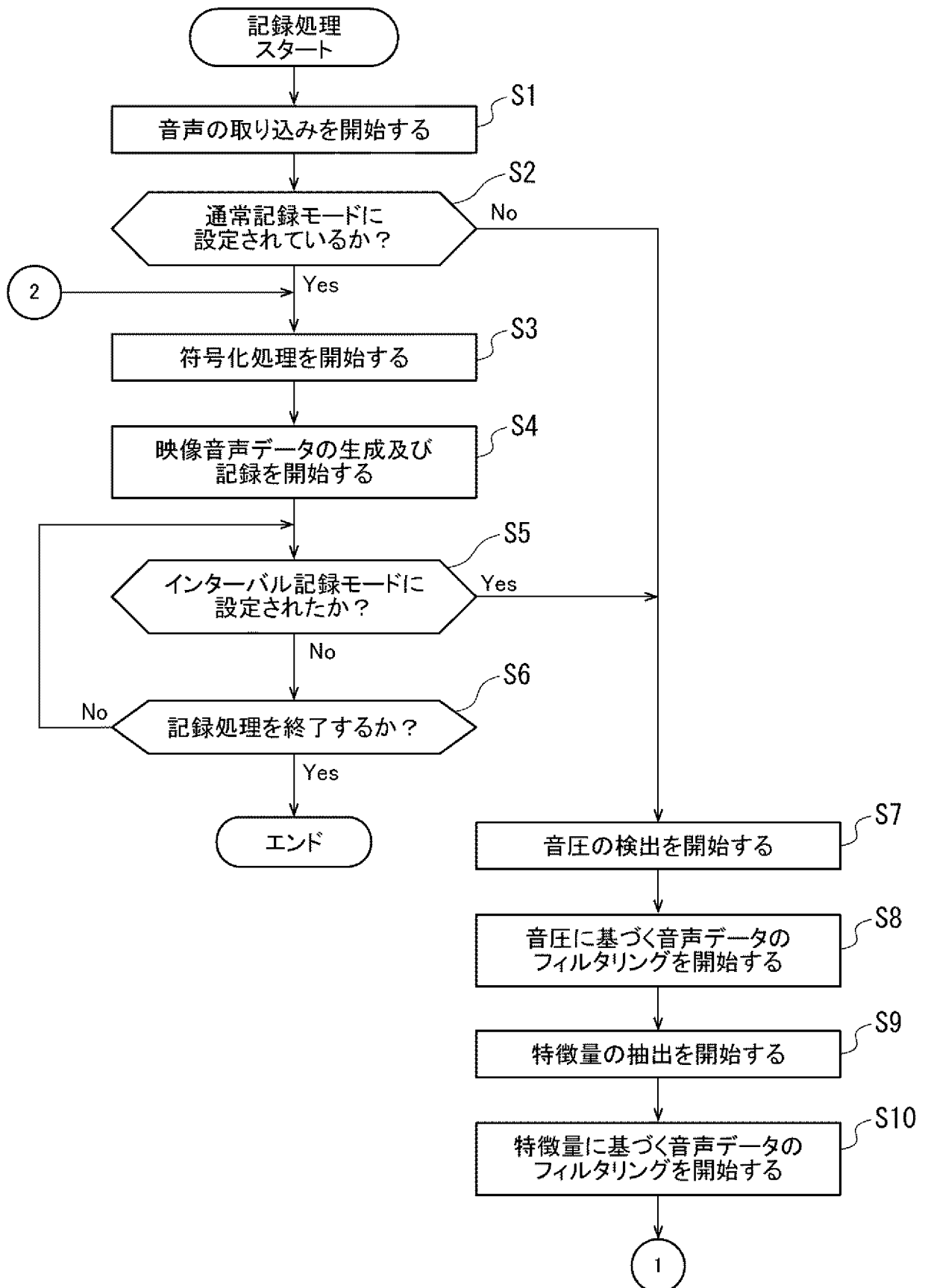


[図3]
FIG. 3

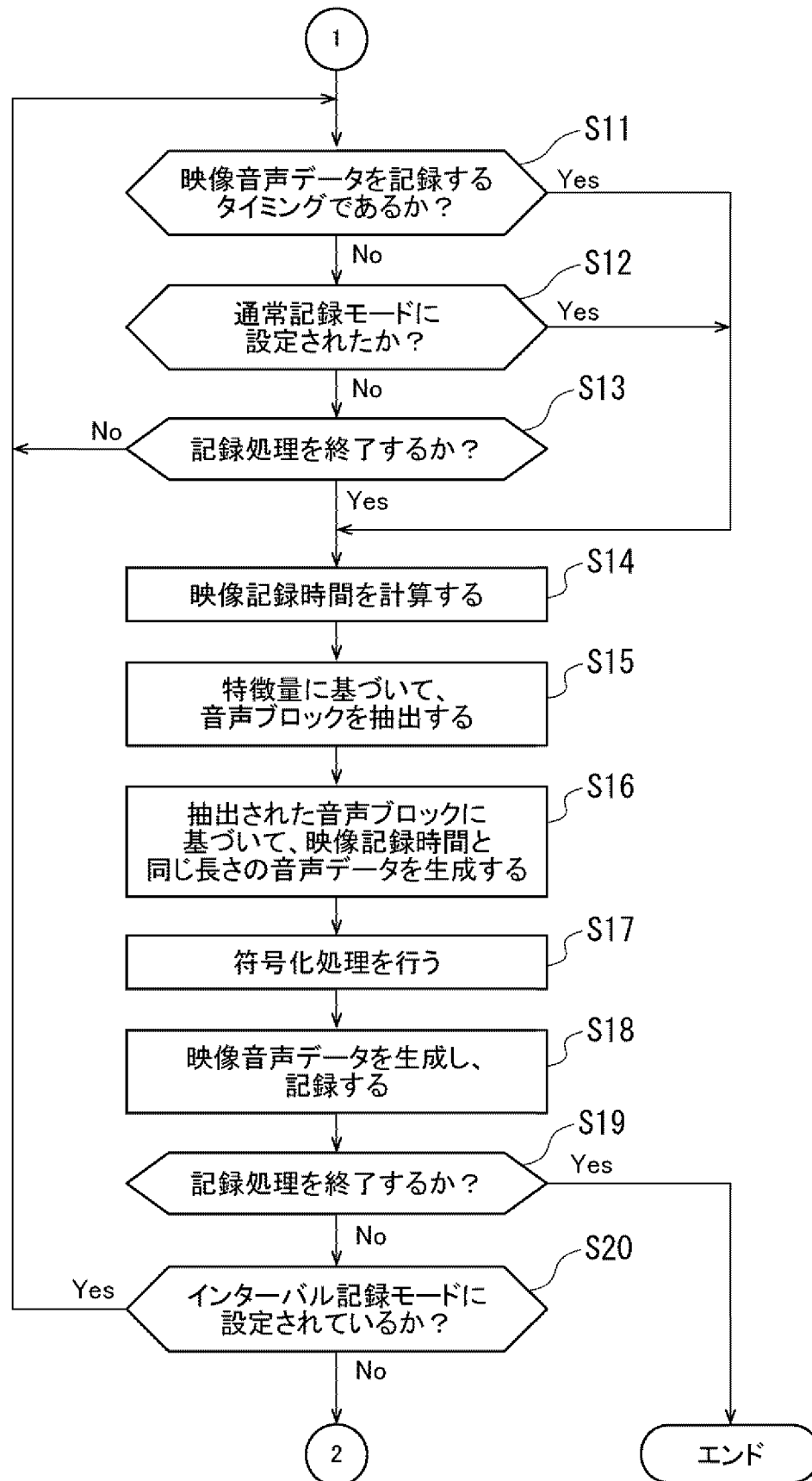


[図4]

FIG. 4

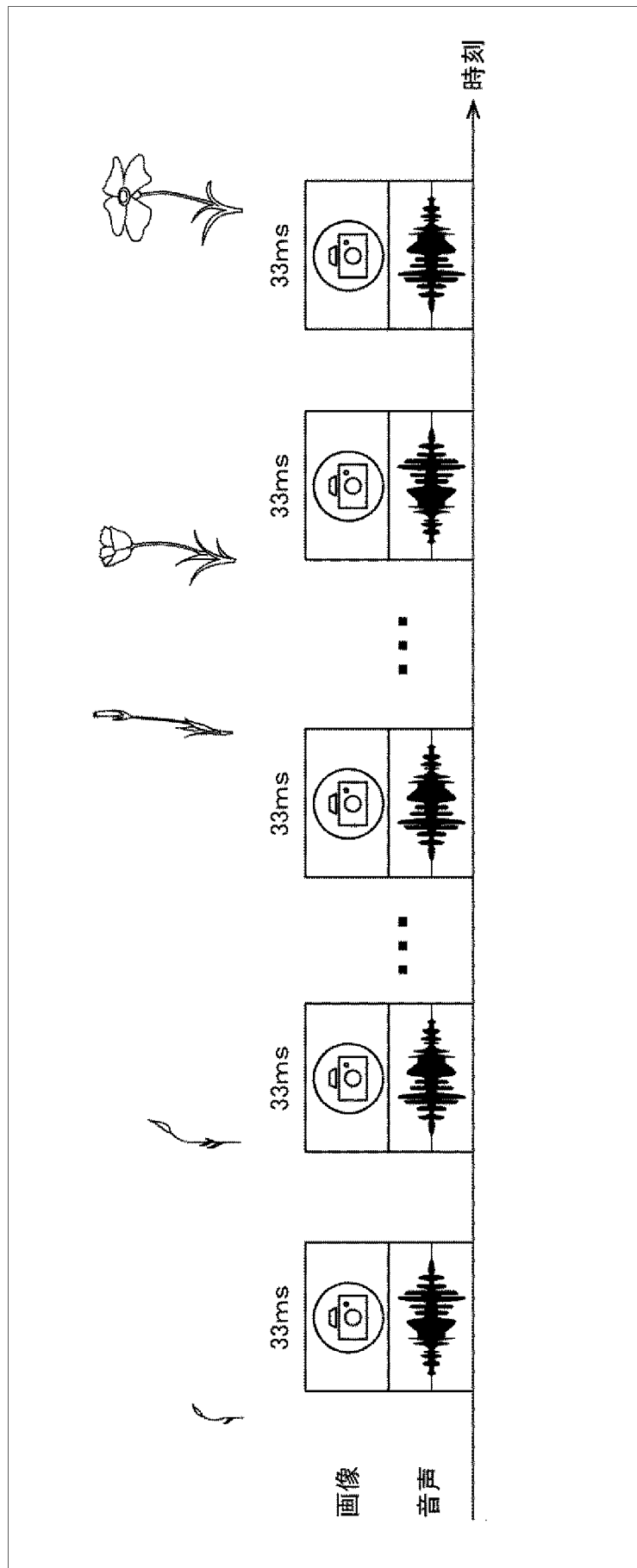


[図5]
FIG. 5

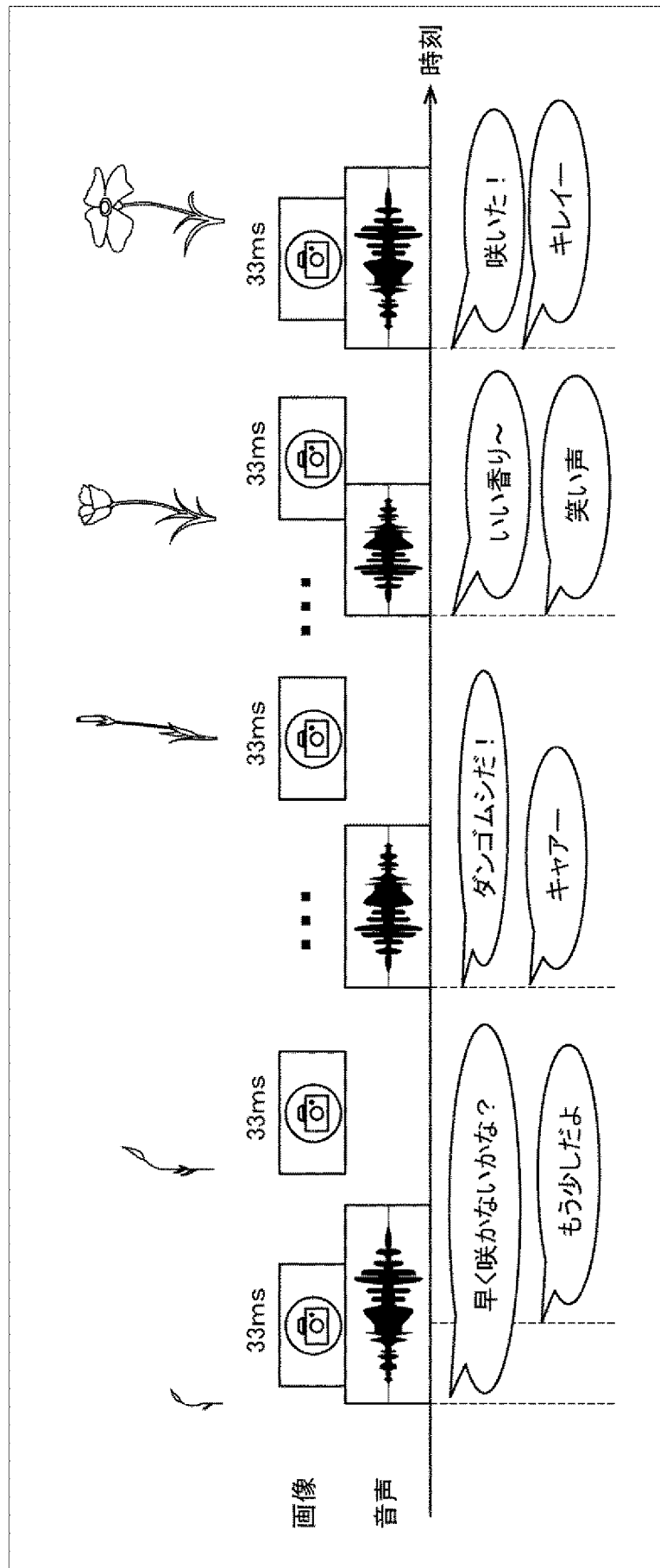


[図6]

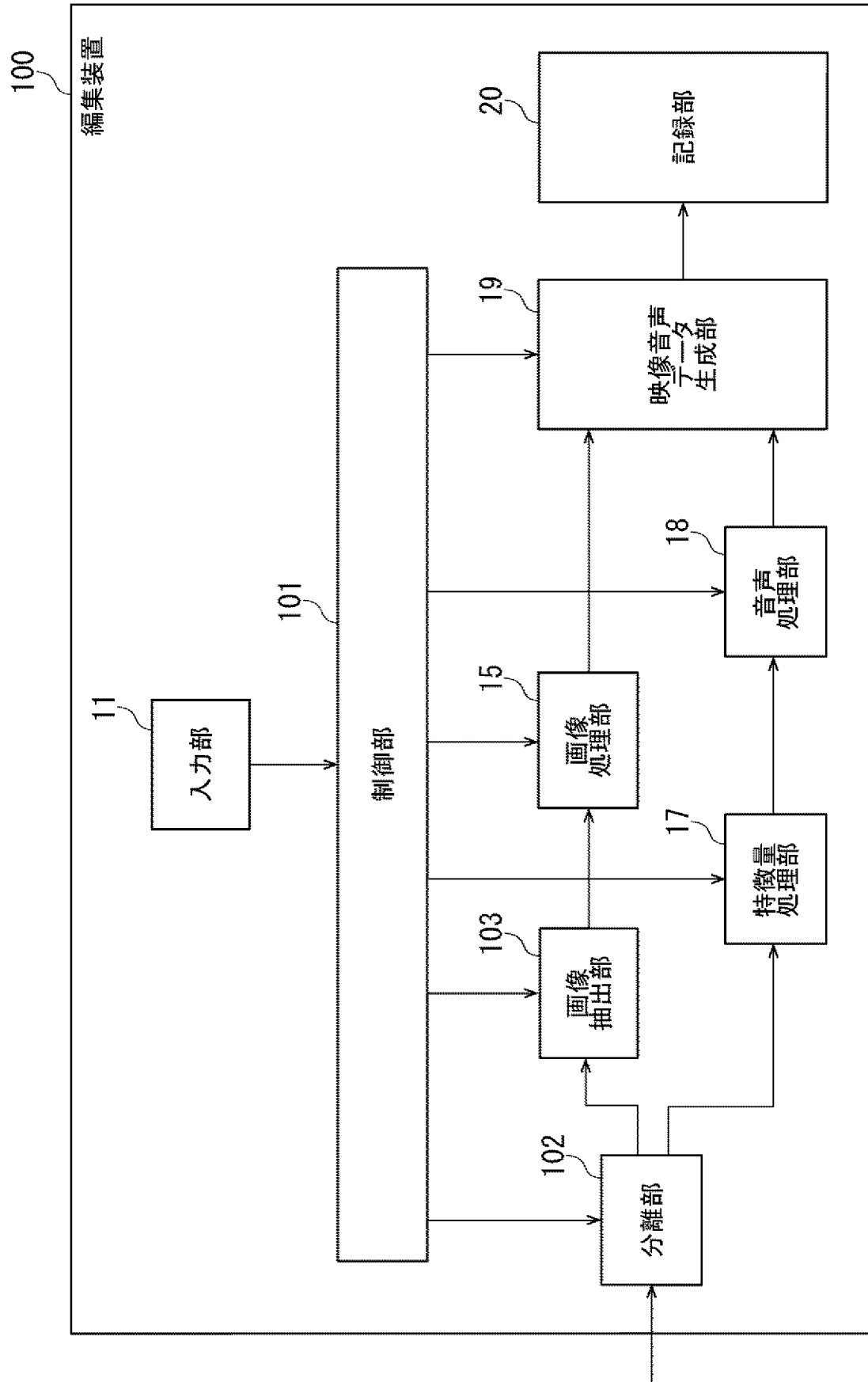
FIG. 6



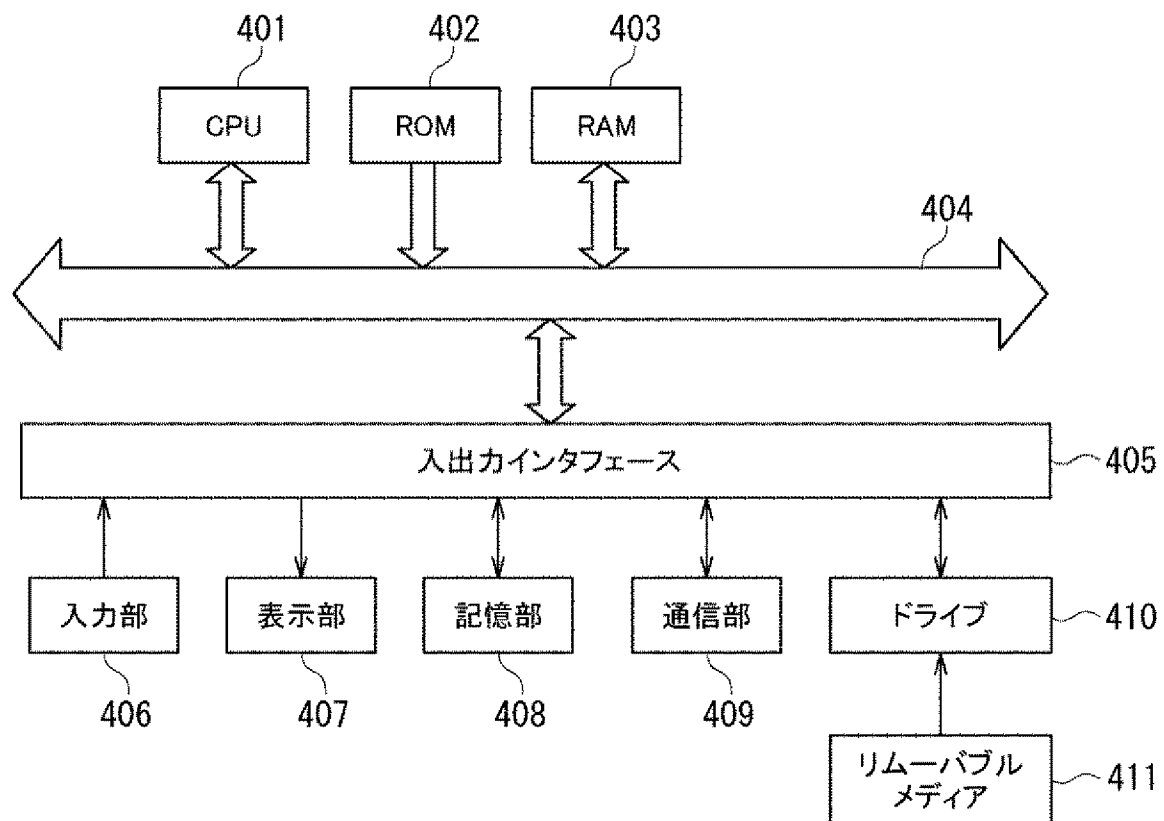
[図7]
FIG. 7



[図8]
FIG. 8



[図9]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/018246

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. H04N5/915(2006.01)i, G10L21/055(2013.01)i, G11B20/10(2006.01)i,
H04N5/77(2006.01)i, H04N5/926(2006.01)i, H04N21/439(2011.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl. H04N5/915, G10L21/055, G11B20/10, H04N5/77, H04N5/926, H04N21/439

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	1971-2018
Registered utility model specifications of Japan	1996-2018
Published registered utility model applications of Japan	1994-2018

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.
X	JP 2016-90774 A (SONY CORP.) 23 May 2016, paragraphs [0020]-[0032], [0239]-[0240], fig. 3 (Family: none)	1-11, 13-17
A	JP 2011-3991 A (CANON INC.) 06 January 2011 (Family: none)	1-17

☐ 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

"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

"I" 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
26 June 2018 (26.06.2018)

Date of mailing of the international search report
10 July 2018 (10.07.2018)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. H04N5/915(2006.01)i, G10L21/055(2013.01)i, G11B20/10(2006.01)i, H04N5/77(2006.01)i, H04N5/926(2006.01)i, H04N21/439(2011.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. H04N5/915, G10L21/055, G11B20/10, H04N5/77, H04N5/926, H04N21/439

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2018年
日本国実用新案登録公報	1996-2018年
日本国登録実用新案公報	1994-2018年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X	JP 2016-90774 A（ソニー株式会社）2016.05.23, 段落[0020]－[0032], [0239]－[0240], 図3（ファミリーなし）	1-11, 13-17
A	JP 2011-3991 A（キヤノン株式会社）2011.01.06, （ファミリーなし）	1-17

☐ C欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

26.06.2018

国際調査報告の発送日

10.07.2018

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

富田 高史

5C

8390

電話番号 03-3581-1101 内線 3541