

(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8月30日(30.08.2018)



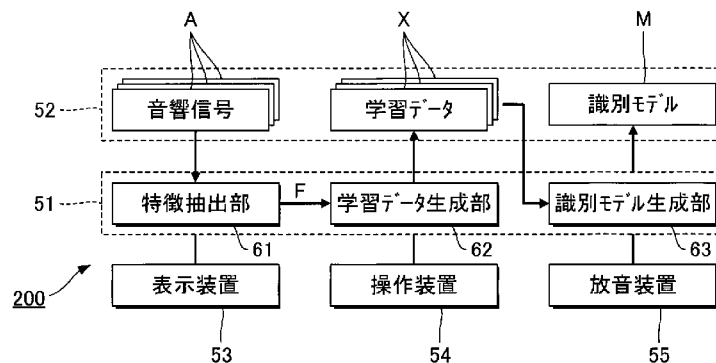
(10) 国際公開番号

WO 2018/155481 A1

- (51) 国際特許分類:
G10L 25/57 (2013.01) *G10L 25/54* (2013.01)
- (21) 国際出願番号: PCT/JP2018/006174
- (22) 国際出願日: 2018年2月21日(21.02.2018)
- (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 2017-035367 2017年2月27日(27.02.2017) JP
- (71) 出願人: ヤマハ株式会社 (YAMAHA CORPORATION) [JP/JP]; 〒4308650 静岡県浜松市中区中沢町10番1号 Shizuoka (JP).
- (72) 発明者: 湯山 雄太(YUYAMA, Yuta); 〒4308650 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内 Shizuoka (JP). 有元 慶太(ARIMOTO, Keita); 〒4308650 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内 Shizuoka (JP).
- (74) 代理人: 大林 章, 外(OHBAYASHI, Akira et al.); 〒1130033 東京都文京区本郷2-15-13 お茶の水ウイングビル6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, 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, KE, KG, KH,

(54) Title: INFORMATION PROCESSING METHOD AND INFORMATION PROCESSING DEVICE

(54) 発明の名称: 情報処理方法および情報処理装置



53 Display device
54 Operation device
55 Sound emission device
61 Feature extraction unit
62 Learning-data generation unit
63 Identification-model generation unit
A Audio signal
M Identification model
X Learning data

(57) Abstract: An information processing device that comprises a learning-data generation unit and an identification-model generation unit. The learning-data generation unit: sets, in accordance with instructions from a user, scene types for any of a plurality of clusters that have been generated by sorting feature values for the sound in one or more pieces of content; and thereby generates a plurality of pieces of learning data that indicate the feature values for the clusters and the scene types that were set for the clusters. On the basis of machine learning that uses the plurality of pieces of learning



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)）

data, the identification-model generation unit generates an identification model that expresses the relationship between feature values and scene types with respect to sound.

(57) 要約：情報処理装置は、1以上のコンテンツにおける音響の特徴量を分類した複数のクラスタの何れかについて利用者からの指示に応じてシーン種別を設定することで、当該クラスタに属する特徴量と、当該クラスタについて設定されたシーン種別とを示す複数の学習データを生成する学習データ生成部と、複数の学習データを利用した機械学習により、音響の特徴量とシーン種別との関係を表す識別モデルを生成する識別モデル生成部とを具備する。

明 細 書

発明の名称： 情報処理方法および情報処理装置

技術分野

[0001] 本発明は、映像および音響を含むコンテンツを解析する技術に関する。

背景技術

[0002] 映像作品等のコンテンツのシーンの種別（以下「シーン種別」という）を解析する各種の技術が従来から提案されている。例えば特許文献１には、コンテンツの映像ストリームから会話シーンまたはレースシーン等の各種のシーン種別を解析する技術が開示されている。

先行技術文献

特許文献

[0003] 特許文献１：特開２００９－２９６２７４号公報

発明の概要

発明が解決しようとする課題

[0004] しかし、コンテンツの映像の傾向とシーン種別との関係は非常に多様であるから、映像ストリームからシーン種別を正確に特定できない場合がある。以上の事情を考慮して、本発明の好適な態様は、コンテンツのシーン種別を適切に特定することを目的とする。

課題を解決するための手段

[0005] 以上の課題を解決するために、本発明の好適な態様に係る情報処理方法は、１以上のコンテンツにおける音響の特徴量を分類した複数のクラスタの何れかについて利用者からの指示に応じてシーン種別を設定することで、当該クラスタに属する特徴量と、当該クラスタについて設定されたシーン種別とを示す複数の学習データを生成する。

また、本発明の他の態様に係る情報処理装置は、１以上のコンテンツにおける音響の特徴量を分類した複数のクラスタの何れかについて利用者からの指示に応じてシーン種別を設定することで、当該クラスタに属する特徴量と

、当該クラスタについて設定されたシーン種別とを示す複数の学習データを生成する学習データ生成部と、前記複数の学習データを利用した機械学習により、音響の特徴量とシーン種別との関係を表す識別モデルを生成する識別モデル生成部とを具備する。

図面の簡単な説明

- [0006] [図1]第1実施形態におけるAVシステムのブロック図である。
[図2]情報処理装置の機能に着目したブロック図である。
[図3]制御装置が実行する処理のフローチャートである。
[図4]第2実施形態における識別モデルの説明図である。
[図5]第3実施形態における情報処理装置のブロック図である。
[図6]第3実施形態の制御装置が実行する処理のフローチャートである。
[図7]第3実施形態における設定画面の模式図である。
[図8]変形例における情報処理装置のブロック図である。

発明を実施するための形態

[0007] <第1実施形態>

図1は、本発明の第1実施形態に係るAV(Audio-Visual)システム10のブロック図である。第1実施形態のAVシステム10は、映像と音響とを含むコンテンツを再生するためのコンピュータシステムである。コンテンツは、例えば映画、スポーツ映像、ライブ映像、ミュージックビデオまたはゲーム等の映像作品である。図1に例示される通り、第1実施形態のAVシステム10は、情報処理装置100と信号供給装置11と放音システム12と表示装置13とを具備する。

- [0008] 信号供給装置11は、映像と音響とを含むコンテンツを表すコンテンツデータDを情報処理装置100に供給する。例えば、可搬型または内蔵型の記録媒体からコンテンツデータDを取得して情報処理装置100に出力する再生機器が信号供給装置11の好適例である。なお、例えば配信サーバ装置(図示略)が送信したコンテンツデータDをインターネット等の通信網から受信する通信装置を、信号供給装置11として利用してもよい。

[0009] 第1実施形態の情報処理装置100は、信号供給装置11から供給されるコンテンツデータDを処理するコンピュータシステム（コンテンツ処理装置）であり、制御装置21と記憶装置22と信号取得部23と再生制御部24とを具備する。なお、相互に別体で用意された複数の装置により情報処理装置100を実現してもよい。

[0010] 信号取得部23は、信号供給装置11からコンテンツデータDを受信するインタフェースである。第1実施形態の信号取得部23は、信号供給装置11から供給されるコンテンツデータDから音響信号A1と映像信号V1とを抽出する。音響信号A1は、コンテンツを構成する音響（例えば映像作品内の会話音、効果音または演奏音等）を表す信号である。第1実施形態の音響信号A1はNチャンネル（例えば5.1チャンネル）で構成される。映像信号V1は、コンテンツを構成する映像を表す信号である。

[0011] 再生制御部24は、音響信号A1および映像信号V1を処理する。具体的には、再生制御部24は、再生制御データQを適用した信号処理により音響信号A1および映像信号V1を調整する。再生制御データQは、コンテンツの再生を制御するためのデータである。図1に例示される通り、第1実施形態の再生制御データQは、音響制御データQaと映像制御データQvとを包含する。音響制御データQaは、コンテンツに含まれる音響の再生を制御するためのデータであり、例えば、コンテンツの音響で形成される音場（例えば仮想音像の定位位置）を制御するための複数のパラメータを含む。他方、映像制御データQvは、コンテンツに含まれる映像の再生を制御するためのデータであり、例えば画像の明度またはコントラスト等の映像特性を制御するための複数のパラメータを含む。

[0012] 図1に例示される通り、第1実施形態の再生制御部24は、音響処理部242と映像処理部244とを含んで構成される。音響処理部242は、信号取得部23から供給される音響信号A1を処理することで音響信号A2を生成する。第1実施形態の音響処理部242は、音響制御データQaを適用した音場制御によりNチャンネルの音響信号A2を生成する。音場制御は、受聴者が

知覚する仮想空間内の残響特性等の音場特性を調整するための音響処理である。例えば、音響処理部 242 は、仮想空間内の複数の音響経路について残響音（初期反射音および後期残響音）を音響制御データ Qa に応じて生成し、複数の残響音を表す信号を音響信号 A1 に混合することで音響信号 A2 を生成する。なお、音響信号 A1 と音響信号 A2 とでチャンネルの総数を相違させてもよい。他方、映像処理部 244 は、信号取得部 23 から供給される映像信号 V1 を処理することで映像信号 V2 を生成する。具体的には、映像処理部 244 は、映像制御データ Qv を適用した画像処理により映像信号 V2 を生成する。

[0013] 放音システム 12 は、例えば N 個のスピーカ装置で構成された N チャンネルのサラウンドシステムであり、音響処理部 242 から供給される N チャンネルの音響信号 A2 が表す音響を放音する。表示装置 13 は、例えば液晶表示パネルで構成され、映像処理部 244 から供給される映像信号 V2 が表す映像を表示する。

[0014] 制御装置 21 は、例えば CPU（Central Processing Unit）等の処理回路を含んで構成され、情報処理装置 100 の各要素を統括的に制御する。記憶装置 22 は、制御装置 21 が実行するプログラムと制御装置 21 が使用する各種のデータとを記憶する。例えば半導体記録媒体および磁気記録媒体等の公知の記録媒体、または複数種の記録媒体の組合せが、記憶装置 22 として任意に採用される。なお、情報処理装置 100 とは別体で記憶装置 22（例えばクラウドストレージ）を用意し、インターネット等の通信網を介して制御装置 21 が記憶装置 22 に対する読出および書込を実行してもよい。

[0015] 図 2 は、制御装置 21 の機能に着目したブロック図である。図 2 に例示される通り、第 1 実施形態の制御装置 21 は、記憶装置 22 に記憶されたプログラムを実行することで、再生制御部 24 が利用する再生制御データ Q を設定するための複数の要素（特徴抽出部 31、解析処理部 32 および設定処理部 33）として機能する。なお、制御装置 21 の一部の機能を専用の電子回路で実現した構成、または、制御装置 21 の機能を複数の装置に分散した構

成も採用される。また、再生制御部 24 の一部または全部の機能を制御装置 21 が実現してもよい。

[0016] 特徴抽出部 31 は、コンテンツの音響に関する特徴量 F を抽出する。具体的には、特徴抽出部 31 は、信号取得部 23 から供給される音響信号 $A1$ を解析することで特徴量 F を抽出する。特徴抽出部 31 による特徴量 F の抽出は、単位期間毎に順次に実行される。単位期間は、所定の時間長（例えば数百ミリ秒）の期間である。

[0017] 第 1 実施形態の特徴量 F は、 N チャンネルにそれぞれ対応する N 個の要素を配列したベクトル（特徴ベクトル）である。特徴量 F のベクトルを構成する任意の 1 個の要素は、音響信号 $A1$ のうち当該要素に対応するチャンネルの周波数特性に関する特徴を表す。具体的には、音響信号 $A1$ の 1 個のチャンネルから抽出される MFCC（Mel-Frequency Cepstrum Coefficients）が、特徴量 F において当該チャンネルに対応する要素として好適である。例えば、特徴量 F は、MFCC を表す U 次元のベクトル（ U 個の係数の系列）を N チャンネルについて配列した、 $(U \times N)$ 個の要素で構成されるベクトルである（ U は自然数）。以上の説明から理解される通り、第 1 実施形態の特徴量 F は、 N チャンネルの音響信号 $A1$ で実現される音場の特徴（すなわち、 N チャンネルの各々における周波数特性の特徴）を表す指標である。

[0018] 解析処理部 32 は、コンテンツにおける音響の特徴量 F に応じて当該コンテンツのシーン種別 S を特定する。シーン種別 S は、コンテンツで表現される場面の分類（カテゴリ）である。例えば、スタンダード、スペクタクル、サイファイ（Sci-Fi）、アドベンチャー、ドラマ、ミュージックビデオ等を含む複数のシーン種別 S の何れかが、解析処理部 32 により特定される。 N チャンネルの音響信号 $A1$ で実現される音場の特徴は、コンテンツのシーン種別 S に依存する傾向がある。例えば、登場人物の台詞が重要な「ドラマ」の場面では、音声に特有の特徴量 F が優勢に観測され、広大な場面を舞台とする「スペクタクル」の場面では、多様な種類の混合音または爆発音等の効果音に特有の周波数特性を N チャンネルの全体が表す複雑な特徴量 F が優勢に

観測される、という傾向がある。以上の傾向を考慮すると、特徴抽出部 3 1 が抽出する特徴量 F は、解析処理部 3 2 がシーン種別 S を特定するための指標として妥当である。解析処理部 3 2 によるシーン種別 S の特定は、特徴抽出部 3 1 による特徴量 F の抽出毎（すなわち単位期間毎）に順次に実行される。以上の説明から理解される通り、特許文献 1 の技術がコンテンツの映像に応じてシーン種別を特定するのとは対照的に、第 1 実施形態の解析処理部 3 2 は、コンテンツの音響に応じてシーン種別 S を特定する。

[0019] 図 2 に例示される通り、記憶装置 2 2 には識別モデル M が記憶される。第 1 実施形態の解析処理部 3 2 は、記憶装置 2 2 に記憶された識別モデル M を利用して、コンテンツのシーン種別 S を特定する。識別モデル M は、音響の特徴量 F とシーン種別 S との関係を表すパターン認識モデルである。第 1 実施形態の識別モデル M は、教師有学習を利用した統計モデルの代表例であるサポートベクターマシン（SVM: Support Vector Machine）である。既存の多数のコンテンツにおける特徴量 F とシーン種別 S との関係に関する傾向を学習するための機械学習により識別モデル M が事前に生成される。具体的には、解析処理部 3 2 は、機械学習で決定された超平面を利用して、コンテンツの音響の特徴量 F が 2 種類のシーン種別 S の何れに該当するかを判別する。2 種類のシーン種別 S の組合せを相違させた複数回にわたり判別を反復することで 1 種類のシーン種別 S が特定される。すなわち、既存の多数のコンテンツにおける特徴量 F とシーン種別 S との関係のもとで、特徴抽出部 3 1 が抽出した特徴量 F に対して最尤である 1 種類のシーン種別 S が推定される。

[0020] 設定処理部 3 3 は、解析処理部 3 2 が特定したシーン種別 S に応じてコンテンツの再生制御データ Q を設定する。図 2 に例示される通り、第 1 実施形態の記憶装置 2 2 は、相異なるシーン種別 S に対応する複数の再生制御データ Q を記憶する。任意の 1 種類のシーン種別 S に対応する再生制御データ Q は、コンテンツの再生に関する複数のパラメータの各々について、当該シーン種別 S に好適な数値を指定する。設定処理部 3 3 は、記憶装置 2 2 に記憶

された複数の再生制御データQのうち解析処理部32が特定したシーン種別Sに対応する再生制御データQを記憶装置22から検索する。そして、設定処理部33は、記憶装置22から検索した再生制御データQ（音響制御データQaおよび映像制御データQv）を再生制御部24に対して指示する。したがって、コンテンツのシーン種別Sに応じた音場特性および映像特性でコンテンツが再生される。第1実施形態では単位期間毎にシーン種別Sが特定されるから、コンテンツのシーン毎に音場特性および映像特性が更新される。

[0021] 図3は、第1実施形態の制御装置21が実行する処理（情報処理方法の例示）を例示するフローチャートである。単位期間毎に図3の処理が実行される。例えば、信号供給装置11によるコンテンツデータDの供給と、放音システム12および表示装置13によるコンテンツの再生とに並行して、図3の処理が実時間的に実行される。

[0022] 図3の処理を開始すると、特徴抽出部31は、信号取得部23から供給される音響信号A1のうち単位期間内の部分から特徴量Fを抽出する（Sa1）。解析処理部32は、特徴抽出部31が抽出した特徴量Fを識別モデルMに付与することで単位期間のシーン種別Sを特定する（Sa2）。

[0023] 以上の手順でシーン種別Sを特定すると、設定処理部33は、当該シーン種別Sに応じてコンテンツの再生制御データQを設定する（Sa3）。具体的には、設定処理部33は、記憶装置22に記憶された複数の再生制御データQのうちシーン種別Sに対応する再生制御データQを検索し、当該再生制御データQ（音響制御データQaおよび映像制御データQv）を再生制御部24に対して指示する。

[0024] 再生制御部24の音響処理部242は、制御装置21が設定した音響制御データQaに応じた音場制御を音響信号A1に対して実行することでNチャンネルの音響信号A2を生成する。音響信号A2は放音システム12に供給される。したがって、コンテンツの音響がシーン種別Sに応じた音場特性で放音システム12から放音される。他方、再生制御部24の映像処理部244は、制御装置21が設定した映像制御データQvに応じた画像処理を映像信号V

1に対して実行することで映像信号V2を生成する。映像信号V2は表示装置13に供給される。したがって、コンテンツの映像がシーン種別Sに応じた映像特性で表示装置13に表示される。

[0025] 以上に説明した通り、第1実施形態では、コンテンツの音響の特徴量Fに応じてシーン種別Sが特定されるから、コンテンツの映像からはシーン種別Sを正確に特定できない場合でも、コンテンツのシーン種別Sを特定することが可能である。第1実施形態では特に、音響の特徴量Fとシーン種別Sとの関係を表す識別モデルMがシーン種別Sの特定に利用される。したがって、特徴量Fとシーン種別Sとの間に想定される傾向のもとで、特徴抽出部31が抽出する特徴量Fにとって最尤であるシーン種別Sを特定することが可能である。

[0026] また、コンテンツの再生を制御するための再生制御データQがシーン種別Sに応じて設定されるから、コンテンツのシーン種別Sにとって適切な特性でコンテンツを再生することが可能である。例えば、第1実施形態では、コンテンツの音響で形成される音場を制御するための音響制御データQaがシーン種別Sに応じて設定される。したがって、コンテンツのシーン種別Sにとって適切な音場でコンテンツの音響を再生することが可能である。

[0027] 第1実施形態では、Nチャンネルの各々の音響信号A1の周波数特性（例えばMFCC）に応じた要素を含むベクトルを特徴量Fとしてシーン種別Sが特定される。すなわち、Nチャンネルの音響信号A1で実現される音場の特徴を表す特徴量Fがシーン種別Sの特定に利用される。したがって、Nチャンネルの音響信号A1で実現される音場の特徴がシーン種別Sに依存するという傾向のもとで、高精度にシーン種別Sを特定することが可能である。

[0028] <第2実施形態>

本発明の第2実施形態を説明する。なお、以下に例示する各形態において、機能または作用が第1実施形態と同様である要素については、第1実施形態の説明で使用了した符号を流用して各々の詳細な説明を適宜に省略する。

[0029] 第1実施形態では、特徴量Fが2種類のシーン種別Sの何れに該当するか

を判別するサポートベクターマシンを識別モデルMとして例示した。第2実施形態では、複数の確率分布の加重和で表現される混合分布を識別モデルM（すなわち混合分布モデル）として利用する。図4は、第2実施形態の識別モデルMの説明図である。なお、特徴量Fが存在し得るベクトル空間として、図4では便宜的に2次元空間が例示されている。図4に例示される通り、第2実施形態の識別モデルMは、相異なるK種類（図4では $K=3$ ）のシーン種別 $S_1 \sim S_K$ にそれぞれ対応するK個の確率分布 $\phi_1 \sim \phi_K$ で構成される混合分布である。例えば各確率分布 ϕ_k （ $k=1 \sim K$ ）を正規分布とした混合正規分布（ガウス混合モデル）が識別モデルMとして好適である。既存の多数のコンテンツにおける特徴量Fとシーン種別 S_k との関係を反映させるための機械学習により識別モデルMが生成される。

[0030] 図4に例示される通り、第2実施形態の解析処理部32は、特徴抽出部31が単位期間毎に抽出した特徴量Fから、K種類のシーン種別 $S_1 \sim S_K$ の各々に対応する尤度 L_k （ $L_1 \sim L_K$ ）を算定する。すなわち、相異なるシーン種別 S_k に対応するK個の尤度 $L_1 \sim L_K$ が単位期間毎に順次に算定される。任意の1種類のシーン種別 S_k に対応する尤度 L_k は、当該シーン種別 S_k の確率分布 ϕ_k のもとで特徴量Fが観測される確度の指標である。すなわち、コンテンツにおいて特徴量Fが抽出された単位期間がシーン種別 S_k に該当する可能性が高いほど、当該シーン種別 S_k に対応する尤度 L_k は大きい数値に設定される。以上の説明から理解される通り、第2実施形態の識別モデルMは、コンテンツの音響の1個の特徴量Fから、K種類のシーン種別 $S_1 \sim S_K$ の各々に関する尤度 L_k を特定可能な統計モデルである。

[0031] 第2実施形態の記憶装置22は、K種類のシーン種別 $S_1 \sim S_K$ にそれぞれ対応するK個の再生制御データ $Q_1 \sim Q_K$ を記憶する。第2実施形態の設定処理部33は、解析処理部32がシーン種別 S_k 毎に算定した尤度 L_k に応じて再生制御データQを設定する。具体的には、設定処理部33は、解析処理部32が算定した尤度 L_k が大きいほど、当該尤度 L_k に対応するシーン種別 S_k の再生制御データ Q_k の内容に近付くように、設定処理部33は再生制御デー

タ Q（音響制御データ Q_aおよび映像制御データ Q_v）を生成する。例えば、各尤度 L_kを加重値とした K 個の音響制御データ Q_aの加重和が音響制御データ Q_aとして音響処理部 242 に指示され、各尤度 L_kを加重値とした K 個の映像制御データ Q_vの加重和が映像制御データ Q_vとして映像処理部 244 に指示される。したがって、K 個の音響制御データ Q_aのうち尤度 L_kが大きいシーン種別 S_kの音響制御データ Q_aほど、放音システム 12 による放音で形成される音場に対して優勢に反映される。また、K 個の映像制御データ Q_vのうち尤度 L_kが大きいシーン種別 S_kの映像制御データ Q_vほど、表示装置 13 に表示される映像に対して優勢に反映される。

[0032] <第 3 実施形態>

図 5 は、第 3 実施形態における情報処理装置 200 のブロック図である。情報処理装置 200 は、第 1 実施形態または第 2 実施形態で利用される識別モデル M を生成するコンピュータシステム（識別モデル生成装置）である。例えば、携帯電話機およびスマートフォン等の可搬型の端末装置またはパーソナルコンピュータ等の情報機器が情報処理装置 200 として利用される。

[0033] 図 5 に例示される通り、情報処理装置 200 は、制御装置 51 と記憶装置 52 と表示装置 53 と操作装置 54 と放音装置 55 とを具備する。表示装置 53 は、例えば液晶表示パネルで構成され、制御装置 51 から指示された画像を表示する。操作装置 54 は、利用者からの操作を受付ける入力機器である。具体的には、利用者が操作可能な複数の操作子、または、表示装置 53 の表示面に対する接触を検知するタッチパネルが、操作装置 54 として好適に利用される。

[0034] 制御装置 51 は、例えば CPU 等の処理回路を含んで構成され、情報処理装置 200 の各要素を統括的に制御する。記憶装置 52 は、制御装置 51 が実行するプログラムと制御装置 51 が使用する各種のデータとを記憶する。例えば半導体記録媒体および磁気記録媒体等の公知の記録媒体、または複数の記録媒体の組合せが、記憶装置 52 として任意に採用される。なお、情報処理装置 100 とは別体で記憶装置 52（例えばクラウドストレージ）を

用意し、インターネット等の通信網を介して制御装置 5 1 が記憶装置 5 2 に対する読出および書込を実行してもよい。第 3 実施形態の記憶装置 5 2 は、多様なシーンを含む複数のコンテンツの各々について N チャンネルの音響信号 A が記憶される。

[0035] 図 5 に例示される通り、第 3 実施形態の制御装置 5 1 は、記憶装置 5 2 に記憶されたプログラムを実行することで、識別モデル M を生成するための複数の要素（特徴抽出部 6 1，学習データ生成部 6 2 および識別モデル生成部 6 3）として機能する。なお、制御装置 5 1 の一部の機能を専用の電子回路で実現した構成、または、制御装置 5 1 の機能を複数の装置に分散した構成も採用される。

[0036] 特徴抽出部 6 1 は、記憶装置 5 2 に記憶された複数の音響信号 A の各々について単位期間毎に特徴量 F を抽出する。特徴抽出部 6 1 が抽出する特徴量 F は、第 1 実施形態と同様に、N チャンネルの各々の周波数特性（例えば M F C C）に応じた N 個の要素で構成される特徴ベクトルである。

[0037] 学習データ生成部 6 2 は、識別モデル M の機械学習に利用される複数の学習データ X を生成する。特徴抽出部 6 1 が抽出した複数の特徴量 F にそれぞれ対応する複数の学習データ X を学習データ生成部 6 2 は生成する。複数の学習データ X の各々は、特徴量 F とシーン種別 S とを含んで構成される。具体的には、任意の 1 個の学習データ X は、音響信号 A から抽出された特徴量 F と、当該音響信号 A に対応するコンテンツのうち当該特徴量 F が抽出された部分の適正なシーン種別 S とを対応させた教師データ（正解データ）である。学習データ X は、特徴量 F に対して正解のシーン種別 S のラベルを付加した教師データとも換言される。第 3 実施形態の学習データ生成部 6 2 は、操作装置 5 4 に対する利用者からの指示に応じて、各学習データ X において特徴量 F に対応させるシーン種別 S を設定する。

[0038] 識別モデル生成部 6 3 は、学習データ生成部 6 2 が生成した複数の学習データ X を利用した機械学習により識別モデル M を生成する。識別モデル生成部 6 3 が生成した識別モデル M が情報処理装置 1 0 0 の記憶装置 2 2 に転送

されて、前述の通り、解析処理部 32 によるシーン種別 S の特定に利用される。識別モデル M は、例えばインターネット等の通信網を介して情報処理装置 100 に転送される。なお、半導体記録媒体等の可搬型の記録媒体を利用して識別モデル M を情報処理装置 200 から情報処理装置 100 に転送してもよい。

[0039] 図 6 は、第 3 実施形態の制御装置 51 が実行する処理（識別モデル生成方法）を例示するフローチャートである。操作装置 54 に対する利用者からの指示を契機として図 6 の処理が開始される。

[0040] 図 6 の処理を開始すると、特徴抽出部 61 は、記憶装置 52 に記憶された複数の音響信号 A の各々から単位期間毎の特徴量 F を抽出する（Sb1）。学習データ生成部 62 は、特徴抽出部 61 が抽出した複数の特徴量 F にそれぞれ対応する複数の学習データ X を生成する（Sb2：学習データ生成処理）。識別モデル生成部 63 は、学習データ生成部 62 が生成した複数の学習データ X を利用した機械学習により識別モデル M を生成して記憶装置 52 に格納する（Sb3）。

[0041] 学習データ生成部 62 が実行する学習データ生成処理 Sb2（学習データ生成方法）の具体的な内容を説明する。学習データ生成処理 Sb2 を開始すると、学習データ生成部 62 は、ステップ Sb1 において特徴抽出部 61 が抽出した複数の特徴量 F を K 個のクラス（部分集合）C に分類する（Sb21）。複数の特徴量 F の分類には、公知のクラス分析技術が任意に採用される。クラスの総数 K は可変値または固定値である。

[0042] 学習データ生成部 62 は、複数の特徴量 F を分類した結果を表す画像（以下「設定画面」という）G を表示装置 53 に表示させる（Sb22）。図 7 は、設定画面 G の模式図である。図 7 に例示される通り、設定画面 G は、特徴量 F の分布を 2 次元空間で表現した画像である。具体的には、設定画面 G は、各特徴量 F を表す特徴画像 P と、各クラス C を表す範囲画像 R（R1, R2, …）とを包含する。同種のシーン種別 S に対応する複数の特徴量の特徴画像 P は、1 個のクラス C に対応する範囲内に局所的に分布する。任意の 1

個のクラスタCに対応する範囲画像Rは、当該クラスタCに分類された複数の特徴量Fの特徴画像Pが分布する領域を包囲する閉図形（例えば楕円形）である。

[0043] 利用者は、操作装置54を適宜に操作することで、設定画面Gにおける任意のクラスタC（具体的には範囲画像R）を選択し、当該クラスタCについて所望のシーン種別Sを指定することが可能である。すなわち、利用者は、任意のクラスタCに属する複数の特徴量Fに対して所望のシーン種別Sをラベリングする。利用者によるシーン種別Sの指示は複数のクラスタCの各々について反復される。以上に例示した利用者からの指示に応じて、学習データ生成部62は、複数のクラスタCの各々にシーン種別Sを設定する（Sb23）。図7に例示される通り、各クラスタCに設定されたシーン種別Sを表す種別画像Z（Z1, Z2, …）が、当該クラスタCの範囲画像Rの近傍に配置される。

[0044] なお、任意の1個のクラスタCを利用者が選択した場合に、当該クラスタCに属する特徴量Fに対応するコンテンツの区間（以下「再生区間」という）を放音装置55により再生してもよい。再生区間は、例えば特徴量Fが抽出された単位期間を含む区間である。具体的には、学習データ生成部62は、音響信号Aのうち特徴量Fが抽出された再生区間を放音装置55に供給することで音響を再生させる。利用者は、放音装置55による再生音を聴取することで、当該再生区間のシーン種別Sを推定できる。任意の1個の特徴量Fの特徴画像Pを利用者が選択した場合に、当該特徴量Fに対応する再生区間を放音装置55に再生させてもよい。なお、各クラスタCのシーン種別Sを利用者に選択させるための構成は任意である。例えば、各クラスタCの条件を対話形式で利用者に順次に回答させることで、各クラスタCのシーン種別Sを設定してもよい。

[0045] 学習データ生成部62は、各クラスタCに対するシーン種別Sを設定した結果に応じて複数の学習データXを生成する（Sb24）。具体的には、特徴抽出部61が抽出した特徴量Fと、当該特徴量Fが属するクラスタCに設定し

たシーン種別Sとを対応させた学習データXが、複数の特徴量Fの各々について生成される。1個のクラスタCに分類された相互に類似する複数の特徴量Fには、共通のシーン種別Sが設定される。以上の説明から理解される通り、第3実施形態の学習データ生成部62は、コンテンツの音響信号Aから抽出された特徴量Fを分類した複数のクラスタCの何れかについて利用者からの指示に応じてシーン種別Sを設定することで、当該クラスタCに属する特徴量Fと、当該クラスタCに設定されたシーン種別Sとを示す学習データXを生成する。

[0046] 以上に説明した通り、第3実施形態では、コンテンツの音響の特徴量Fとコンテンツのシーン種別Sとを示す複数の学習データXが生成される。したがって、コンテンツの映像からはシーン種別Sを正確に特定できない場合でも、学習データXを利用した機械学習で生成された識別モデルMにより、当該コンテンツのシーン種別Sを特定することが可能である。また、各クラスタCに属する特徴量Fと、当該クラスタCについて利用者からの指示に応じて設定されたシーン種別Sとを示す複数の学習データXが生成されるから、利用者の意図または嗜好を反映した識別モデルMを生成できるという利点もある。

[0047] <変形例>

以上に例示した各態様は多様に変形されてもよい。具体的な変形の態様を以下に例示する。以下の例示から任意に選択された2個以上の態様を、相互に矛盾しない範囲で適宜に併合してもよい。

[0048] (1) 単位期間毎にシーン種別Sを特定する前述の各形態では、短時間で頻繁にシーン種別S（さらには再生制御データQ）が変化する可能性がある。以上の事情を考慮すると、シーン種別Sの頻繁な変化を抑制するための構成が好適である。例えば、複数の単位期間の各々について、当該単位期間と前後の他の単位期間とを含む所定個の単位期間のなかで最頻のシーン種別Sを選択してもよい。以上の構成によれば、瞬間的なシーン種別Sの変化を抑制することが可能である。

- [0049] (2) 前述の各形態では、コンテンツの音響および映像の双方に着目したが、映像に関する要素は省略されてもよい。例えば、第1実施形態または第2実施形態において、映像処理部244（さらには映像制御データQv）および表示装置13を省略してもよい。
- [0050] (3) 前述の各形態では、音響信号A1の周波数特性に応じた要素をNチャンネルについて配列した特徴ベクトルを特徴量Fとして例示したが、特徴量Fの種類は以上の例示に限定されない。例えば、相異なる音階音（例えば平均律の12半音の各々）に対応する複数の要素を含むクロマベクトル（PCP: Pitch Class Profile）を特徴量Fとして利用してもよい。クロマベクトルのうち任意の1個の音階音に対応する要素は、音響信号A1のうち当該音階音に対応する帯域成分の信号強度を複数のオクターブにわたり加算または平均した数値に設定される。また、音響信号A1の信号強度（レベル）に応じた要素をNチャンネルについて配列したN次元ベクトルを特徴量Fとして利用してもよい。なお、音響信号A1の全部のチャンネルを特徴量Fの抽出に利用する必要はない。例えば、Nチャンネルの音響信号A1の何れかから特徴量Fを抽出してもよい。
- [0051] (4) 第1実施形態ではサポートベクターマシンを識別モデルMとして例示し、第2実施形態では混合分布モデル（例えばガウス混合モデル）を識別モデルMとして例示したが、識別モデルMの種類は以上の例示に限定されない。例えば、ニューラルネットワークまたはベイズ分類等の各種のパターン認識モデルを識別モデルMとして利用される。以上の例示から理解される通り、識別モデルMは、音響の特徴量Fとシーン種別Sとの関係を表す統計モデルとして包括的に表現される。具体的には、識別モデルMは、既存の多数のコンテンツにおける特徴量Fとシーン種別Sとの間で成立する関係のもとで、当該識別モデルMに付与された特徴量Fに対して妥当性の高いシーン種別Sを判別する仕組みである。
- [0052] (5) 移動体通信網またはインターネット等の通信網を介して端末装置（例えば携帯電話機またはスマートフォン）と通信するサーバ装置に、第1実施

形態または第2実施形態に係る情報処理装置100の少なくとも一部の機能を実行させてもよい。

[0053] 例えば、図8に例示される通り、端末装置300と通信可能な情報処理装置100（例えばウェブサーバ）に特徴抽出部31と解析処理部32と設定処理部33とが設置される。特徴抽出部31は、端末装置300から通信網400を介して音響信号A1を受信し、当該音響信号A1から特徴量Fを抽出する（Sa1）。解析処理部32は、前述の各形態と同様に、特徴量Fからシーン種別Sを特定する（Sa2）。設定処理部33は、シーン種別Sに応じた再生制御データQを設定し、当該再生制御データQを端末装置300に送信する（Sa3）。再生制御データQを利用した処理を実行する再生制御部24は、端末装置300に搭載される。

[0054] なお、特徴抽出部31を端末装置300に設置し、特徴抽出部31が抽出した特徴量Fを端末装置300から情報処理装置100に送信してもよい。情報処理装置100からは、特徴抽出部31が省略される。解析処理部32は、端末装置300から受信した特徴量Fからシーン種別Sを特定する。また、設定処理部33を端末装置300に設置してもよい。情報処理装置100からは、設定処理部33が省略される。解析処理部32は、特徴量Fから特定したシーン種別Sを端末装置300の設定処理部33に通知する。また、再生制御部24を情報処理装置100に設置し、再生制御部24による処理後の音響信号A2および映像信号V2を端末装置300に送信してもよい。

[0055] （6）第3実施形態に係る情報処理装置200の少なくとも一部の機能を、情報処理装置200とは別体の外部装置で実現してもよい。例えば、携帯電話機またはスマートフォン等の端末装置で情報処理装置200を実現する構成のもとで、端末装置と通信可能なサーバ装置に特徴抽出部61が設置される。すなわち、情報処理装置200から特徴抽出部61は省略される。また、識別モデルMを生成する第3実施形態の機能を、第1実施形態または第2実施形態の情報処理装置100に搭載してもよい。

[0056] （7）前述の各形態では、コンテンツの音響の特徴量Fからシーン種別Sを

特定したが、コンテンツの音響の特徴量Fと当該コンテンツの映像の特徴量とからシーン種別Sを特定してもよい。すなわち、音響の特徴量Fに加えて映像の特徴量も加味してシーン種別Sを特定する構成は、本発明の範囲から除外されない。

[0057] (8) 第1実施形態または第2実施形態に係る情報処理装置100は、前述の例示の通り、制御装置21とプログラムとの協働により実現される。第1実施形態および第2実施形態から把握されるプログラムは、コンピュータに、映像と音響とを含むコンテンツにおける音響の特徴量Fに応じて当該コンテンツのシーン種別Sを特定する解析処理(Sa2)を実行させる。さらに好適な態様に係るプログラムは、コンピュータに、解析処理で特定したシーン種別Sに応じて、コンテンツの再生を制御するための再生制御データQを設定する設定処理(Sa3)を実行させる。

[0058] また、第3実施形態に係る情報処理装置200は、前述の例示の通り、制御装置51とプログラムとの協働により実現される。第3実施形態から把握されるプログラムは、1以上のコンテンツにおける音響の特徴量Fを分類した複数のクラスタCの何れかについて利用者からの指示に応じてシーン種別Sを設定することで、当該クラスタCに属する特徴量Fと、当該クラスタCについて設定されたシーン種別Sとを示す学習データXを生成する学習データ生成処理(Sb2)と、複数の学習データXを利用した機械学習により、音響の特徴量Fとシーン種別Sとの関係を表す識別モデルMを生成する識別モデル生成処理(Sb3)とを実行させる。

[0059] 以上に例示した各態様のプログラムは、コンピュータが読取可能な記録媒体に格納された形態で提供されてコンピュータにインストールされる。記録媒体は、例えば非一過性(non-transitory)の記録媒体であり、CD-ROM等の光学式記録媒体(光ディスク)が好例であるが、半導体記録媒体または磁気記録媒体等の公知の任意の形式の記録媒体も包含される。なお、非一過性の記録媒体とは、一過性の伝搬信号(transitory, propagating signal)を除く任意の記録媒体を含み、揮発性の記録媒体を除外するものではない。

また、通信網を介した配信の形態でプログラムをコンピュータに提供してもよい。

[0060] (9) 以上に例示した形態（特に第1実施形態および第2実施形態）から、例えば以下の態様が把握される。

[0061] 本発明の好適な態様（態様A1）に係る情報処理方法は、映像と音響とを含むコンテンツにおける前記音響の特徴量に応じて当該コンテンツのシーン種別を特定する。以上の態様では、コンテンツにおける音響の特徴量に応じてシーン種別が特定されるから、コンテンツの映像からはシーン種別を正確に特定できない場合でも、当該コンテンツのシーン種別を特定することが可能である。

[0062] 態様A1の好適例（態様A2）では、前記シーン種別の特定において、音響の特徴量とシーン種別との関係を表す識別モデルを利用して、前記コンテンツにおける前記音響の特徴量から当該コンテンツのシーン種別を特定する。以上の態様では、音響の特徴量とシーン種別との関係を表す識別モデルがシーン種別の特定に利用される。したがって、音響の特徴量とシーン種別との間に想定される傾向のもとで、特徴抽出部が抽出する特徴量に対して妥当性が高いシーン種別を特定することが可能である。

[0063] 態様A2の好適例（態様A3）において、前記識別モデルは、前記コンテンツにおける前記音響の特徴量からひとつのシーン種別を特定する統計モデルである。以上の態様によれば、コンテンツの音響の特徴量からひとつのシーン種別を特定することが可能である。

[0064] 態様A2の好適例（態様A4）において、前記識別モデルは、前記コンテンツにおける前記音響の特徴量から、複数のシーン種別の各々に関する尤度を特定する統計モデルである。以上の態様によれば、コンテンツの音響の特徴量から、複数のシーン種別の各々に関する尤度を特定することが可能である。

[0065] 態様A1から態様A4の何れかの好適例（態様A5）において、前記コンテンツの再生を制御するための再生制御データを、前記特定したシーン種別

に応じて設定する。以上の態様では、コンテンツの再生を制御するための再生制御データがシーン種別に応じて設定される。したがって、コンテンツのシーン種別にとって適切な特性でコンテンツを再生することが可能である。

[0066] 態様 A 5 の好適例（態様 A 6）において、前記再生制御データは、前記コンテンツの音響で形成される音場を制御するための音響制御データを含む。以上の態様では、コンテンツの音響で形成される音場を制御するための音響制御データがシーン種別に応じて設定される。したがって、コンテンツのシーン種別にとって適切な音場でコンテンツの音響を再生することが可能である。

[0067] 本発明の好適な態様（態様 A 7）に係る情報処理方法は、映像と音響とを含むコンテンツにおける前記音響の特徴量から特定された当該コンテンツのシーン種別に応じて、前記コンテンツの再生を制御するための再生制御データを設定する。以上の態様では、コンテンツの音響の特徴量から特定されたシーン種別に応じて、コンテンツの再生を制御するための再生制御データが設定される。したがって、コンテンツの映像からはシーン種別を正確に特定できない場合でも、当該コンテンツのシーン種別を特定し、当該シーン種別にとって適切な特性でコンテンツを再生することが可能である。

[0068] 態様 A 1 から態様 A 7 の何れかの好適例（態様 A 8）において、前記特徴量は、前記コンテンツの音響を表す複数のチャンネルにそれぞれ対応する複数の要素を含むベクトルであり、前記複数の要素の各々は、当該要素に対応するチャンネルの周波数特性に応じた数値である。以上の態様では、複数のチャンネルの各々の周波数特性に応じた要素を含むベクトルを特徴量としてシーン種別が特定される。すなわち、複数のチャンネルの音響信号で実現される音場の特徴を表す特徴量がシーン種別の特定に利用される。したがって、複数のチャンネルの音響信号で実現される音場の特徴がシーン種別に依存するという傾向のもとで、コンテンツのシーン種別を高精度に特定することが可能である。

[0069] 本発明の好適な態様（態様 A 9）に係る情報処理装置は、映像と音響とを

含むコンテンツにおける前記音響の特徴量に応じて当該コンテンツのシーン種別を特定する解析処理部を具備する。以上の態様では、コンテンツにおける音響の特徴量に応じてシーン種別が特定されるから、コンテンツの映像からはシーン種別を正確に特定できない場合でも、当該コンテンツのシーン種別を特定することが可能である。

[0070] 態様A 9の好適例（態様A 10）に係る情報処理装置は、前記解析処理部が特定したシーン種別に応じて、前記コンテンツの再生を制御するための再生制御データを設定する設定処理部と、前記設定処理部が設定した再生制御データに応じて前記コンテンツを再生する再生制御部とを具備する。以上の態様では、コンテンツの再生を制御するための再生制御データがシーン種別に応じて設定される。したがって、コンテンツのシーン種別にとって適切な特性でコンテンツを再生することが可能である。

[0071] （10）以上に例示した形態（特に第3実施形態）から、例えば以下の態様が把握される。

[0072] 本発明の好適な態様（態様B 1）に係る情報処理方法は、1以上のコンテンツにおける音響の特徴量を分類した複数のクラスタの何れかについて利用者からの指示に応じてシーン種別を設定することで、当該クラスタに属する特徴量と、当該クラスタについて設定されたシーン種別とを示す複数の学習データを生成する。以上の態様では、コンテンツの音響の特徴量とコンテンツのシーン種別とを示す学習データが生成される。したがって、複数の学習データを利用した機械学習で生成される識別モデルを利用することで、コンテンツの映像からはシーン種別を正確に特定できない場合でも、当該コンテンツのシーン種別を特定することが可能である。また、1個のクラスタに属する特徴量と、当該クラスタについて利用者からの指示に応じて設定されたシーン種別とを示す複数の学習データが生成されるから、利用者の意図または嗜好を反映した識別モデルを生成できるという利点もある。

[0073] 態様B 1の好適例（態様B 2）において、前記複数の学習データを利用した機械学習により、音響の特徴量とシーン種別との関係を表す識別モデルを

生成する。以上の態様では、コンテンツの音響の特徴量とコンテンツのシーン種別とを示す複数の学習データが生成され、複数の学習データを利用した機械学習で識別モデルが生成される。この識別モデルを利用することで、コンテンツの映像からはシーン種別を正確に特定できない場合でも、当該コンテンツのシーン種別を特定することが可能である。

[0074] 態様B2の好適例（態様B3）において、前記識別モデルは、前記コンテンツにおける前記音響の特徴量からひとつのシーン種別を特定する統計モデルである。以上の態様によれば、コンテンツの音響の特徴量からひとつのシーン種別を特定することが可能である。

[0075] 態様B2の好適例（態様B4）において、前記識別モデルは、前記コンテンツにおける前記音響の特徴量から、複数のシーン種別の各々に関する尤度を特定する統計モデルである。以上の態様によれば、コンテンツの音響の特徴量から、複数のシーン種別の各々に関する尤度を特定することが可能である。

[0076] 態様B1から態様B4の何れかの好適例（態様B5）において、前記シーン種別は、前記コンテンツで表現される場面の分類である。

[0077] 本発明の好適な態様（態様B6）に係る情報処理装置は、1以上のコンテンツにおける音響の特徴量を分類した複数のクラスタの何れかについて利用者からの指示に応じてシーン種別を設定することで、当該クラスタに属する特徴量と、当該クラスタについて設定されたシーン種別とを示す複数の学習データを生成する学習データ生成部と、前記複数の学習データを利用した機械学習により、音響の特徴量とシーン種別との関係を表す識別モデルを生成する識別モデル生成部とを具備する。以上の態様では、コンテンツの音響の特徴量とコンテンツのシーン種別とを示す学習データが生成され、複数の学習データを利用した機械学習で識別モデルが生成される。したがって、この識別モデルを利用することで、コンテンツの映像からはシーン種別を正確に特定できない場合でも、当該コンテンツのシーン種別を特定することが可能である。また、1個のクラスタに属する特徴量と、当該クラスタについて利

用者からの指示に応じて設定されたシーン種別とを示す複数の学習データが生成されるから、利用者の意図または嗜好を反映した識別モデルを生成できるという利点もある。

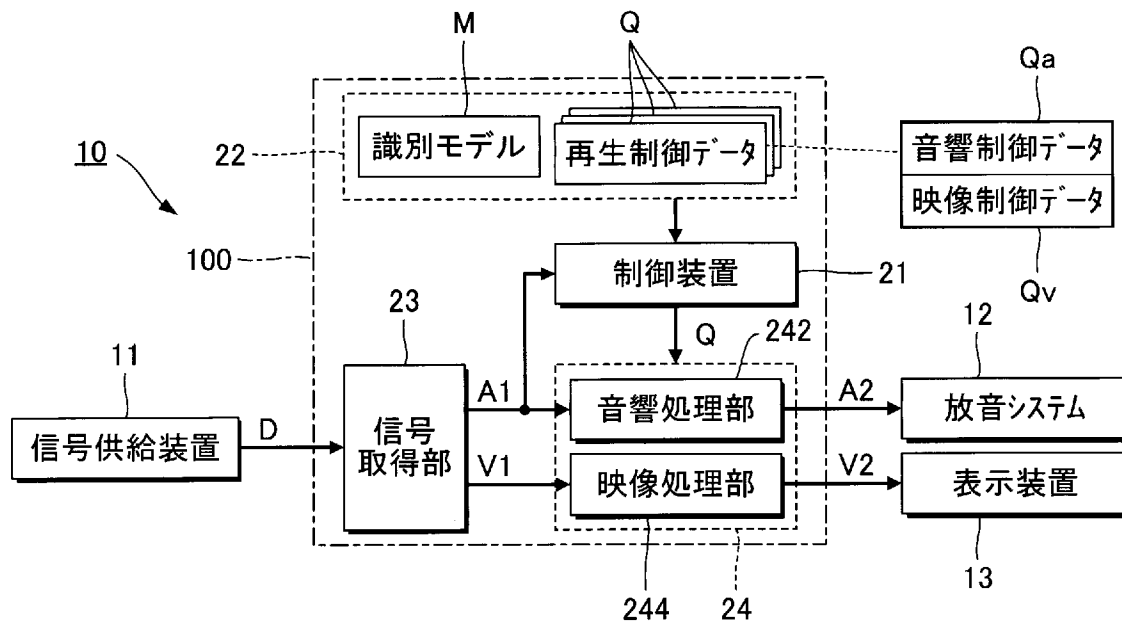
符号の説明

[0078] 10…AVシステム、100、200…情報処理装置、300…端末装置、400…通信網、11…信号供給装置、12…放音システム、13…表示装置、21…制御装置、22…記憶装置、23…信号取得部、24…再生制御部、242…音響処理部、244…映像処理部、31…特徴抽出部、32…解析処理部、33…設定処理部、51…制御装置、52…記憶装置、53…表示装置、54…操作装置、55…放音装置、61…特徴抽出部、62…学習データ生成部、63…識別モデル生成部。

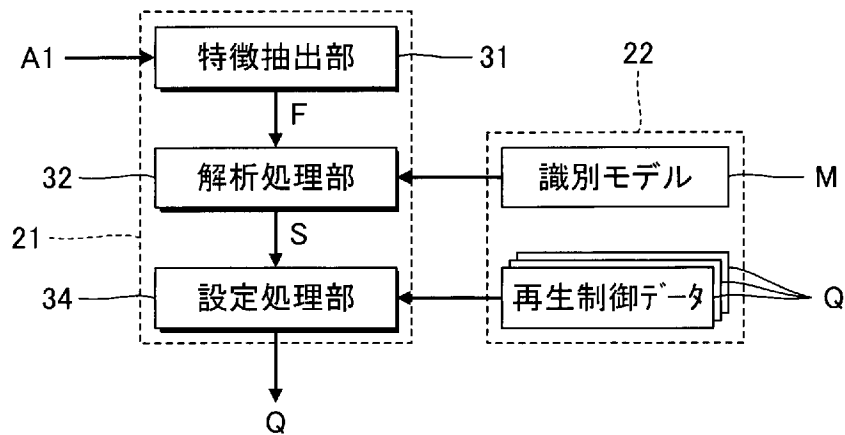
請求の範囲

- [請求項1] 1以上のコンテンツにおける音響の特徴量を分類した複数のクラスタの何れかについて利用者からの指示に応じてシーン種別を設定することで、当該クラスタに属する特徴量と、当該クラスタについて設定されたシーン種別とを示す複数の学習データを生成する、
コンピュータにより実現される情報処理方法。
- [請求項2] 前記複数の学習データを利用した機械学習により、音響の特徴量とシーン種別との関係を表す識別モデルを生成する
請求項1の情報処理方法。
- [請求項3] 前記識別モデルは、前記コンテンツにおける前記音響の特徴量からひとつのシーン種別を特定する統計モデルである
請求項2の情報処理方法。
- [請求項4] 前記識別モデルは、前記コンテンツにおける前記音響の特徴量から、複数のシーン種別の各々に関する尤度を特定する統計モデルである
請求項2の情報処理方法。
- [請求項5] 前記シーン種別は、前記コンテンツで表現される場面の分類である
請求項1から請求項4の何れかの情報処理方法。
- [請求項6] 1以上のコンテンツにおける音響の特徴量を分類した複数のクラスタの何れかについて利用者からの指示に応じてシーン種別を設定することで、当該クラスタに属する特徴量と、当該クラスタについて設定されたシーン種別とを示す複数の学習データを生成する学習データ生成部と、
前記複数の学習データを利用した機械学習により、音響の特徴量とシーン種別との関係を表す識別モデルを生成する識別モデル生成部とを具備する情報処理装置。

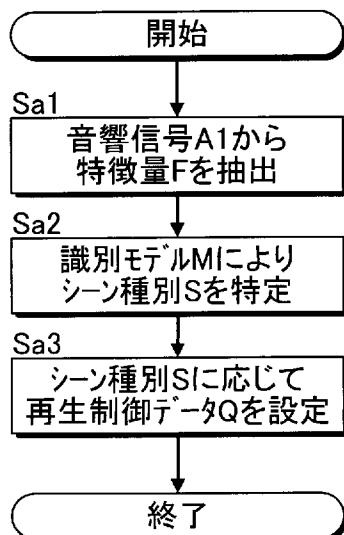
[図1]



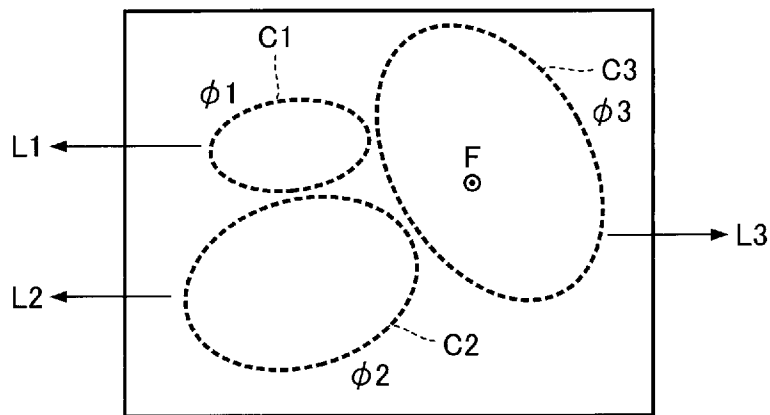
[図2]



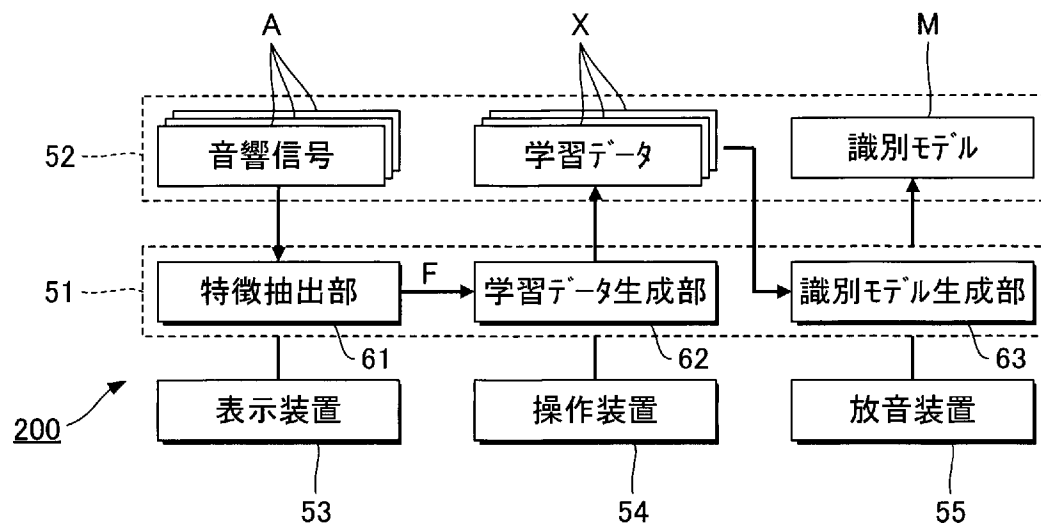
[図3]



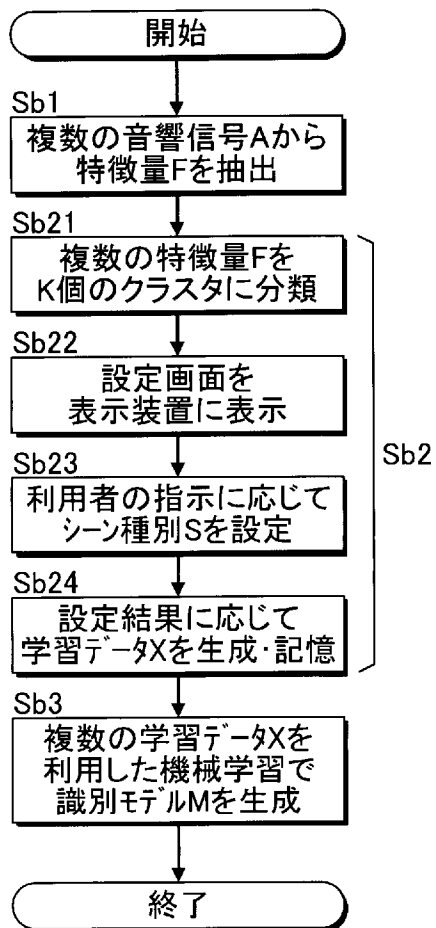
[図4]



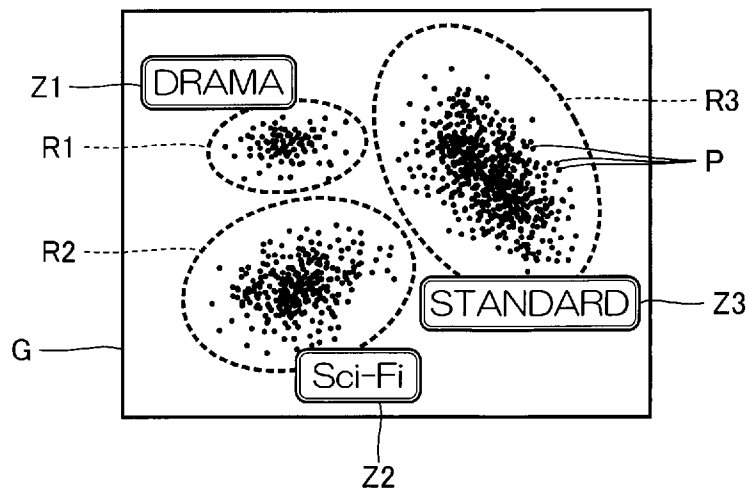
[図5]



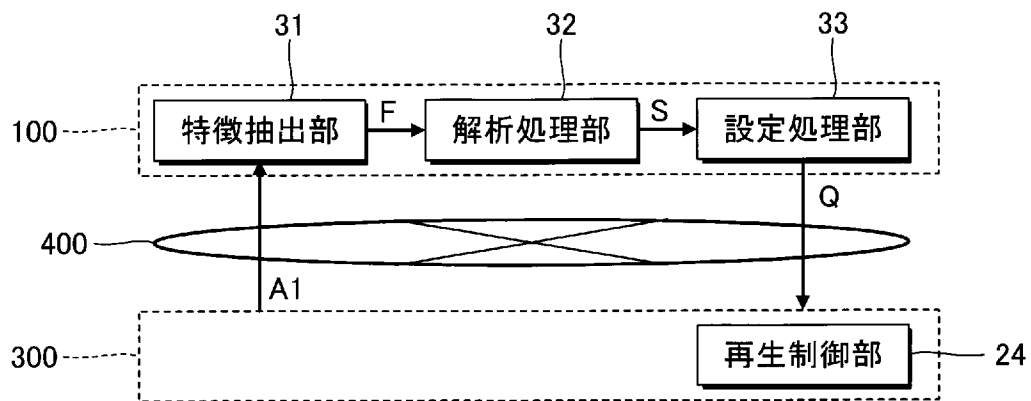
[図6]



[図7]



[図8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/006174

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G10L25/57 (2013.01) i, G10L25/54 (2013.01) i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl. G10L25/57, G10L25/54

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 2011-95680 A (NEC CORPORATION) 12 May 2011, paragraphs [0026]-[0028], [0042], fig. 5 (Family: none)	1-6
A	KEMP, T. et al., Strategies for automatic segmentation of audio data, Proc. ICASSP'00, 05 June 2000, pp. 1423-1426, particularly, chapter 4	1-6



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

"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
12.04.2018

Date of mailing of the international search report
01.05.2018

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

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2018/006174

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2010-38943 A (TOSHIBA CORPORATION), 18 February 2010, entire text, all drawings (Family: none)	1-6

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G10L25/57(2013.01)i, G10L25/54(2013.01)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. G10L25/57, G10L25/54

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 2011-95680 A (日本電気株式会社) 2011.05.12, 段落[0026]-[0028], [0042], 第5図 (ファミリーなし)	1-6
A	KEMP, Thomas, et al., Strategies for Automatic Segmentation of Audio Data, Proc. ICASSP '00, 2000.06.05, pp.1423-1426, 特に 第4章	1-6
A	JP 2010-38943 A (株式会社東芝)	1-6

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

12.04.2018

国際調査報告の発送日

01.05.2018

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)

間宮 嘉誉

電話番号 03-3581-1101 内線 3591

5Z

3454

C (続き) . 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
	2010.02.18, 全文, 全図 (ファミリーなし)	