

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7207836号
(P7207836)

(45)発行日 令和5年1月18日(2023.1.18)

(24)登録日 令和5年1月10日(2023.1.10)

(51)国際特許分類 F I
H 0 4 N 21/442 (2011.01) H 0 4 N 21/442
H 0 4 N 17/00 (2006.01) H 0 4 N 17/00 M

請求項の数 18 外国語出願 (全50頁)

(21)出願番号	特願2021-135077(P2021-135077)	(73)特許権者	518239651
(22)出願日	令和3年8月20日(2021.8.20)		テレビジョン・インサイツ、インコーポ レイテッド
(62)分割の表示	特願2018-536185(P2018-536185)の分割		アメリカ合衆国 0 2 1 3 9 マサチュー セッツ州、ケンブリッジ、マサチューセ ッツ アベニュー # 1 5 7、9 5 5
原出願日	平成29年1月6日(2017.1.6)	(74)代理人	110000877
(65)公開番号	特開2021-184644(P2021-184644 A)		弁理士法人R Y U K A国際特許事務所
(43)公開日	令和3年12月2日(2021.12.2)	(72)発明者	シドウ、インダービル
審査請求日	令和3年8月23日(2021.8.23)		アメリカ合衆国、マサチューセッツ州 0 2 4 2 1、レキシントン、ワルサム・ ストリート 4 4 7
(31)優先権主張番号	62/275,699	(72)発明者	ヤンフェン、リュ
(32)優先日	平成28年1月6日(2016.1.6)		アメリカ合衆国、ニューヨーク州 1 1 1 0 6、ロング・アイランド・シティ、 最終頁に続く
(33)優先権主張国・地域又は機関	米国(US)		

(54)【発明の名称】 視聴者エンゲージメントを評価するためのシステム

(57)【特許請求の範囲】

【請求項1】

セットトップボックスまたはケーブル接続に接続されたテレビジョン上で再生しているビデオとの視聴者エンゲージメントを評価するためのシステムであって、

前記テレビジョンの前の視聴エリアを撮像するように配設された、前記視聴エリアの画像データを取得するための少なくとも1つのカメラと、

前記テレビジョンに近接して配設された、前記テレビジョンに結合されたスピーカによって発せられたオーディオデータを取得するためのマイクロフォンと、

前記少なくとも1つのカメラおよび前記マイクロフォンに動作可能に結合された、プロセッサ実行可能命令を記憶し、前記画像データをサードパーティに送信する可能性をなくすために、前記画像データおよび前記オーディオデータが前記少なくとも1つのカメラからの後続の画像データおよび前記マイクロフォンからの後続のオーディオデータで上書きされる前に100ミリ秒以下の間、前記画像データおよび前記オーディオデータを記憶するためのメモリと、

前記少なくとも1つのカメラ、前記マイクロフォン、および前記メモリに動作可能に結合され、前記テレビジョン、前記セットトップボックス、またはケーブル接続への接続を有さないプロセッサと、ここにおいて、前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記少なくとも1つのカメラからの前記画像データ、および前記マイクロフォンからの前記オーディオデータを受信することと、

前記オーディオデータに少なくとも一部基づいて、テレビジョン上に表示された前記ビデオの識別情報を決定することと、

前記画像データおよび前記画像データに現れる顔の識別情報に少なくとも一部基づいて、前記視聴エリア中に存在する人々を識別するためのコンピュータビジョンモデルをトレーニングすることと、

前記コンピュータビジョンモデルを使用して、前記後続の画像データに基づいて前記視聴エリアにいる各人物についての一意の識別子を決定することと、

前記後続の画像データに少なくとも一部基づいて、前記視聴エリア中に存在する人々の第1の数、および前記視聴エリア中の前記ビデオにエンゲージしている人々の第2の数を推定することと、

前記ビデオについての注意レートを推定し、前記注意レートに基づいて注意インデックスを決定することによって、前記ビデオの前記視聴者エンゲージメントを評価することであって、前記注意レートは、所与の時間期間にわたる、前記視聴エリア中の人々の前記第1の数に対する、前記ビデオにエンゲージしている人々の前記第2の数の比を表し、前記注意インデックスは、前記注意レートの平均として定義される、評価することと

を行い、

前記プロセッサに動作可能に結合された、利用可能なアップストリーム帯域幅または送信されるデータの量のうちの少なくとも1つに基づいて選択された送信ウィンドウ中にバースト送信で、人々の前記第1の数、人々の前記第2の数、前記ビデオの前記識別情報、および前記視聴エリア中の前記人々のための前記一意の識別子をリモートサーバに送信するためのネットワークインターフェースと、

を備える、システム。

【請求項2】

前記少なくとも1つのカメラは、実質的に毎秒1フレーム以上のフレームレートで前記画像データを取得する、請求項1に記載のシステム。

【請求項3】

前記プロセッサは、身体骨格検出に基づいて、前記視聴エリア中に存在する人々の前記第1の数を推定する、請求項1又は2に記載のシステム。

【請求項4】

前記プロセッサは、視線追跡に基づいて、前記ビデオにエンゲージしている人々の前記第2の数を推定する、請求項1から3のいずれか一項に記載のシステム。

【請求項5】

前記マイクロフォンは、約0.1Hzの取得レートで前記オーディオデータを取得する、請求項1から4の何れか一項に記載のシステム。

【請求項6】

前記プロセッサは、オーディオ信号フィンガープリンティングを使用して、前記ビデオの前記識別情報を決定する、請求項1から5のいずれか一項に記載のシステム。

【請求項7】

前記ビデオは、複数のビデオ中の一意のビデオであり、前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記画像データと前記視聴エリア中に潜在的に存在する人々に関する人口統計学的情報とに基づいて、視聴者カウントとポジティブ持続時間比とを推定し、前記視聴者カウントは、各一意のビデオにエンゲージしている人々の前記第2の数を表し、前記ポジティブ持続時間比は、前記一意のビデオの持続時間に対する、前記視聴エリア中の人々が前記一意のビデオを見ることによって費やされた合計時間の比を表す、請求項1から6のいずれか一項に記載のシステム。

【請求項8】

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記リモートサーバへの送信のために、人々の前記第1の数および人々の前記第2の数を前記メモリに記憶する、

10

20

30

40

50

請求項 1 から 7 のいずれか一項に記載のシステム。

【請求項 9】

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、
前記視聴エリア中に存在する各人物の感情を推定する

請求項 1 から 8 のいずれか一項に記載のシステム。

【請求項 10】

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記オーディオデータに少なくとも一部基づいて、複数のビデオ中のあらかじめ決定されたビデオが前記テレビジョン上に表示されているかどうかを決定し、ここにおいて、前記視聴者エンゲージメントを評価することは、前記あらかじめ決定されたビデオが表示されているかどうか少なくとも一部基づく、

請求項 1 から 9 のいずれか一項に記載のシステム。

【請求項 11】

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記画像データから前記視聴エリア中の各人物についての人口統計学的情報を推定する、請求項 1 から 10 のいずれか一項に記載のシステム。

【請求項 12】

前記人口統計学的情報は、年齢、性別、民族グループ、および顔の表情を含む、請求項 11 に記載のシステム。

【請求項 13】

前記ビデオはストリーミングコンテンツを含む、請求項 1 から 12 のいずれか一項に記載のシステム。

【請求項 14】

セットトップボックスを介してテレビジョン上で再生しているテレビ番組との視聴者エンゲージメントを評価するためのシステムであって、

前記テレビジョンの前の視聴エリアを撮像するように配設された、実質的に毎秒 1 フレーム以上のフレームレートで前記視聴エリアの画像データを取得するための少なくとも 1 つのカメラと、

前記テレビジョンに近接して配設された、約 0.1 Hz の取得レートで前記テレビジョンによって発せられたオーディオデータを取得するためのマイクロフォンと、

前記少なくとも 1 つのカメラおよび前記マイクロフォンに動作可能に結合された、プロセッサ実行可能命令を記憶するためのメモリと、

前記少なくとも 1 つのカメラ、前記マイクロフォン、および前記メモリに動作可能に結合され、前記テレビジョンにも前記セットトップボックスにも接続されないプロセッサと、ここにおいて、前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記少なくとも 1 つのカメラからの前記画像データ、および前記マイクロフォンからの前記オーディオデータを受信することと、

オーディオ信号フィンガープリンティングを使用して前記オーディオデータに少なくとも一部基づいて、前記テレビジョン上に表示された前記テレビ番組の識別情報を決定することと、

前記画像データに少なくとも一部基づいて、前記視聴エリア中に存在する人々の第 1 の数を推定することと、

前記視聴エリア中の前記テレビ番組にエンゲージしている人々の第 2 の数を推定することと、

前記テレビ番組についての注意レートを推定し、前記注意レートに基づいて注意インデックスを決定することによって、前記テレビ番組の前記視聴者エンゲージメントを評価することであって、前記注意レートは、所与の時間期間にわたる、前記視聴エリア中の人々の前記第 1 の数に対する、前記テレビ番組にエンゲージしている人々の前記第 2 の数の比を表し、前記注意インデックスは、前記注意レートの平均として定義される、評価することと

10

20

30

40

50

を行い、

前記プロセッサに動作可能に結合された、人々の前記第 1 の数、人々の前記第 2 の数、および前記テレビ番組の前記識別情報をリモートサーバに送信するためのネットワークインターフェースと、
を備える、システム。

【請求項 15】

前記テレビ番組は、複数のテレビ番組中の一意のテレビ番組であり、前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記画像データと前記視聴エリア中に潜在的に存在する人々に関する人口統計学的情報とに基づいて、視聴者カウントとポジティブ持続時間比とを推定し、前記視聴者カウントは、各一意のテレビ番組にエンゲージしている人々の前記第 2 の数を表し、前記ポジティブ持続時間比は、前記一意のテレビ番組の持続時間に対する、前記視聴エリア中の人々が前記一意のテレビ番組を見ることによって費やされた合計時間の比を表す、
請求項 14 に記載のシステム。

10

【請求項 16】

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記画像データに少なくとも一部基づいて、前記視聴エリア中に存在する各人物の識別情報を決定することと、

識別された各人物についての前記視聴者エンゲージメントを評価することと
を行う、請求項 14 または 15 に記載のシステム。

20

【請求項 17】

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記オーディオデータに少なくとも一部基づいて、複数のテレビ番組中のあらかじめ決定されたテレビ番組が前記テレビジョン上に表示されているかどうかを決定し、ここにおいて、前記視聴者エンゲージメントを評価することは、前記あらかじめ決定されたテレビ番組が表示されているかどうか少なくとも一部基づく、
請求項 14 から 16 のいずれか一項に記載のシステム。

【請求項 18】

前記プロセッサに動作可能に結合された、人物によって装着されたワイヤレス信号送信機からのワイヤレス信号を検出するための少なくとも 1 つのワイヤレス信号受信機をさらに備え、

前記プロセッサは、前記ワイヤレス信号に基づいて、前記人物と前記テレビジョンとの間の距離および前記テレビジョンに対する前記人物のロケーションを測定する、
請求項 14 から 17 のいずれか一項に記載のシステム。

30

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

[0001]本出願は、その全体が参照により本明細書に組み込まれる、2016年1月6日に出願された「SYSTEMS AND METHODS FOR ASSESSING VIEWER ENGAGEMENT」と題する米国出願第62/275,699号の優先権を主張する。

40

【背景技術】

【0002】

[0002]TVオーディエンス測定(TV audience measurement)の従来の方法は、オーディエンスからデータを収集するために、ピープルメータ(people meter)および日記を使用することを含む。これらの方法は、一般に、TV受像機が設置された部屋にいる人間(潜在的オーディエンスメンバ)を認識しようとする。該方法はまた、TVで再生している一連の画像(たとえば、TV番組またはコマーシャル広告)をキャプチャすることを伴い得る。次いで、各画像について、特定の画像が表示されたときに部屋にいる人々の

50

数が推定され得る。

【0003】

[0003]これらの方法はいくつかの欠陥を有する。第一に、これらの方法によって収集されたデータは、通常、TVが設置された部屋にいる人々の数のみを含む。データは、一般に、視聴者がどのくらいの頻度でTVを実際に見ているかについて示さない（測定はTVがオンのときに行われる）。第二に、収集されたデータは、人々がどのくらいの頻度で特定のチャンネルに合わせるかについて示し得る。ただし、それは、番組または広告に対する人々の反応を測定せず、したがって、番組または広告の有効性について示さない。第三に、TVレーティングは、世帯またはコミュニティにおける特定の層（demographics）について与えられない。

10

【発明の概要】

【0004】

[0004]本発明の実施形態は、TVオーディエンスの視聴者エンゲージメント（viewer engagement）を評価する装置、システム、および方法を含む。一例では、ディスプレイ上で再生しているビデオとの視聴者エンゲージメントを定量化するためのシステムは、ディスプレイの前の視聴エリア（viewing area）を撮像するように配設された、視聴エリアの画像データを取得するための少なくとも1つのカメラを含む。ディスプレイに結合されたスピーカによって発せられたオーディオデータを取得するために、マイクロフォンがディスプレイに近接して配設される。本システムはまた、カメラおよびマイクロフォンに動作可能に結合された、プロセッサ実行可能命令を記憶するためのメモリと、カメラ、マイク

20

【0005】

[0005]別の例では、ディスプレイ上に示されるビデオとの視聴者エンゲージメントを定量化する方法は、少なくとも1つのカメラを用いて、ビデオがディスプレイ上に示されている間、ディスプレイの前の視聴エリアの画像を取得することを含む。本方法はまた、マイクロフォンを用いて、ディスプレイに結合されたスピーカによって発せられたビデオのサウンドトラックを表すオーディオデータを取得することを含む。本方法は、カメラおよびプロセッサに動作可能に結合されたプロセッサを用いて、オーディオデータに少なくとも一部基づいて、ビデオの識別情報を決定することと、プロセッサを用いて、画像データに少なくとも一部基づいて、ビデオがディスプレイ上に示されている間の視聴エリア中に存在する人々の第1の数、および視聴エリア中のビデオにエンゲージしている人々の第2の数を推定することとをさらに含む。本方法はまた、プロセッサによって、ビデオの識別情報、人々の第1の数、および人々の第2の数をリモートサーバに送信することを含む。

30

40

【0006】

[0006]また別の例では、ディスプレイ上で再生しているビデオとの視聴者エンゲージメントを評価するためのシステムが開示される。ディスプレイは、ビデオのサウンドトラックを発するスピーカに結合される。本システムは、ビデオがディスプレイ上で再生している間、第1のサンプルレートで、ディスプレイの前の視聴エリアの可視画像を取得するための可視カメラを含む。第1のサンプルレートで、ビデオがディスプレイ上で再生している間、ディスプレイの前の視聴エリアの赤外線画像を取得するために、赤外線カメラが本システムに含まれる。第1のサンプルレートよりも低い第2のサンプルレートで、ビデオがディスプレイ上で再生している間、スピーカによって発せられたサウンドトラックのサンプルを取得するために、マイクロフォンがディスプレイに近接して配設される。本シス

50

テムはまた、可視カメラ、赤外線カメラ、およびマイクロフォンに動作可能に結合された、(i) サウンドトラックのサンプルに基づいてビデオを識別することと、(i i) 可視画像および赤外線画像に基づいて、ビデオがディスプレイ上で再生している間の視聴エリア中の人々の数、およびビデオにエンゲージしている人々の数を推定することと、(i i i) サウンドトラックのサンプル、可視画像、および赤外線画像を上書き、消去、および/または廃棄することとを行うためのプロセッサを含む。本システムはまた、プロセッサに動作可能に結合された、ビデオの識別情報、ビデオがディスプレイ上で再生している間の視聴エリア中の人々の数、およびビデオにエンゲージしている人々の数の表現 (representation) を記憶するためのメモリを含む。本システムは、プロセッサに動作可能に結合された、表現をサーバに送信するためのネットワークインターフェースをさらに含む。

10

【 0 0 0 7 】

[0007]また別の例では、複数のビデオ中の一意のビデオについての視聴者エンゲージメントを定量化する方法は、複数の世帯中の各世帯において、ディスプレイの前の視聴エリアの画像データを取得することと、ディスプレイが複数のビデオ中のビデオを示しているかどうかを決定することとを含む。本方法はまた、複数のビデオ中の各一意のビデオについて、画像データおよび複数の世帯中の各世帯に関する人口統計学的情報 (demographic information) に基づいて、(i) 視聴レート (viewing rate) および (i i) ウォッチングレート (watching rate) を推定することを含む。視聴レートは、ビデオを示しているディスプレイの合計数に対する、視聴エリア中の人々の合計数の比を表し、ウォッチングレートは、複数の世帯中の人々の合計数に対する、ディスプレイがビデオを示している世帯中の人々の合計数の比を表す。本方法はまた、複数のビデオ中の各一意のビデオについて、視聴レートおよびウォッチングレートに基づいて、視聴可能性インデックス (viewability index) を決定することを含む。

20

【 0 0 0 8 】

[0008]上記の概念と以下でさらに詳細に説明される追加の概念のすべての組合せが、(そのような概念が相互に矛盾しないという条件で) 本明細書で開示される本発明の主題の一部であるものとして企図されることを諒解されたい。特に、本開示の最後に記載される、請求する主題のすべての組合せが、本明細書で開示される本発明の主題の一部であるものとして企図される。また、参照により組み込まれる開示にも記載され得る、本明細書で明示的に採用された用語は、本明細書で開示される特定の概念に最も一致する意味を与えられるべきであることを諒解されたい。

30

【 0 0 0 9 】

[0009]図面は、主に説明のためであり、本明細書で説明される本発明の主題の範囲を限定するものではないことを、当業者なら理解するであろう。図面は、必ずしも一定の縮尺であるとは限らず、いくつかの事例では、本明細書で開示される本発明の主題の様々な態様は、異なる特徴の理解を容易にするために、図面において誇張または拡大されて示されることがある。図面では、同様の参照符号は、概して、同様の特徴 (たとえば、機能的に類似する要素および/または構造的に類似する要素) を指す。

【 図面の簡単な説明 】

【 0 0 1 0 】

40

【 図 1 】 [0010] 図 1 は、TV オーディエンスの視聴者エンゲージメントを評価するためのシステムの概略図を示す。

【 図 2 A 】 [0011] 図 2 A は、図 1 に示されているシステムを使用してユーザエンゲージメントを定量化する方法を示す図である。

【 図 2 B 】 [0012] 図 2 B は、ユーザエンゲージメントを定量化するためのコンピュータビジョンモデルをトレーニングする方法を示す図である。

【 図 3 A 】 [0013] 図 3 A は、顔および眼球追跡、顔認識、ならびに感情分析を含む、視聴者エンゲージメントの方法を示す図である。

【 図 3 B 】 [0014] 図 3 B は、視聴可能性インデックスおよび注意インデックス (attention index) の概念を示す図である。

50

【図 4 A】[0015]図 4 A は、視聴可能性インデックスを推定することを含む、視聴者エンゲージメントを評価するためのプロセスを示す図である。

【図 4 B】[0016]図 4 B は、注意インデックスを推定することを含む、視聴者エンゲージメントを評価するためのプロセスを示す図である。

【図 5】[0017]図 5 は、視聴エリア中の各人物の顔の向き (orientation) を決定することを含む、視聴者エンゲージメントを評価するためのプロセスを示す図である。

【図 6】[0018]図 6 は、骨格、顔、識別情報、感情、およびエンゲージメントを検出するためのプロセスを示す図である。

【図 7】[0019]図 7 は、視聴者エンゲージメント評価の例示的な方法におけるデータ取得アーキテクチャの概略図である。

10

【図 8 A】[0020]図 8 A は、図 7 に示されているアーキテクチャを使用して取得されたコマmercialメッセージ (CM) 曲線を示す図である。

【図 8 B】図 8 B は、図 7 に示されているアーキテクチャを使用して取得されたコマmercialメッセージ (CM) 曲線を示す図である。

【図 8 C】図 8 C は、図 7 に示されているアーキテクチャを使用して取得されたコマmercialメッセージ (CM) 曲線を示す図である。

【図 8 D】図 8 D は、図 7 に示されているアーキテクチャを使用して取得されたコマmercialメッセージ (CM) 曲線を示す図である。

【図 8 E】図 8 E は、図 7 に示されているアーキテクチャを使用して取得されたコマmercialメッセージ (CM) 曲線を示す図である。

20

【図 8 F】図 8 F は、図 7 に示されているアーキテクチャを使用して取得されたコマmercialメッセージ (CM) 曲線を示す図である。

【図 8 G】図 8 G は、図 7 に示されているアーキテクチャを使用して取得されたコマmercialメッセージ (CM) 曲線を示す図である。

【図 9】[0021]図 9 は、サンプリングされた TV 局の各々についての CM 曲線の比を示す図である。

【図 10】[0022]図 10 は、表 5 に示されている決定木の決定結果をもつ、決定木による分類モデルを示す図である。

【図 11】[0023]図 11 は、CM の長さに対する視聴可能性レートを示す図である。

【図 12】[0024]図 12 は、番組の開始からの経過時間と視聴可能性レートとの間の相関を示す図である。

30

【図 13】[0025]図 13 は、図 1 ~ 図 12 に示されている技術を使用して取得された視聴者エンゲージメントデータの通信を示す図である。

【図 14】[0026]図 14 は、図 1 ~ 図 12 に示されている技術を使用して取得された視聴者エンゲージメントデータの配布および使用を示す図である。

【図 15】[0027]図 15 は、図 1 ~ 図 12 に示されている技術を使用して取得された視聴者エンゲージメントデータのビッグデータ分析および視覚化を示す図である。

【図 16】[0028]図 16 は、図 1 ~ 図 12 に示されている技術を使用して取得された視聴者エンゲージメントデータを補完するための追加のデータを取得するためのモデルを示す図である。

40

【発明を実施するための形態】

【0011】

[0029] TV オーディエンス測定の従来の方法における欠点に対処するために、本明細書で開示されるシステムおよび方法は、ビデオ (たとえば、TV ショー、映画、ウェブショー、広告、または他のコンテンツ) を再生しているディスプレイ (たとえば、TV、コンピュータ、またはタブレット) の前の視聴エリアの画像データを取得する。例示的なシステムは、画像データから、何人の人々が視聴エリアにいるか、およびそれらの人々のうちの誰がビデオを実際に見ているかを決定する。本システムはまた、マイクロフォンを用いてビデオのサウンドトラックをサンプリングし、サウンドトラックのサンプルを使用してビデオを識別する。本システムは、ビデオ、視聴エリア中の人々の数、およびビデオを見

50

ている人々の数に関する情報を、ローカルメモリに記憶（および/または残存）し、その情報をインターネットまたは他のネットワーク接続を介してリモートサーバに送信する。

【0012】

[0030]ビデオ自体に埋め込まれたデジタルウォーターマークに基づいてビデオを識別する、ビデオとの視聴者エンゲージメントを測定するための以前のシステムとは異なり、本発明のシステムの例は、ビデオのサウンドトラックに基づいてビデオを識別する。したがって、本発明のシステムは、視聴者の構内にあるディスプレイ、セットトップボックス、またはケーブル接続に接続される必要はない。このことは、それらを取り付けるおよび取り外すことを容易にする（したがって、採用される可能性が高くなる）。このことはまた、それらが誤動作する可能性、またはディスプレイがオフであるのにセットトップボックスをオンのままにすることによって引き起こされる「フォールスポジティブ」インプレッションを記録する可能性を低くする。

10

【0013】

[0031]本発明のシステムはまた、視聴エリア中の人々の数とビデオにエンゲージしている人々の数とを決定するために、ローカルで、すなわち、視聴者の構内で画像データを処理する。それはまた、視聴エリアに誰かがいる間に表示されているビデオを識別するために、ローカルでオーディオデータを処理することができる。それは、ローカルに、すなわち、視聴者の構内のローカルデバイス中のまたはそれに結合されたメモリに、このデータを記憶する。処理された画像およびオーディオデータは、未処理（raw）の画像およびオーディオデータよりもはるかに少ないメモリを消費し、したがって、このローカルメモリは、より長い時間期間をカバーする情報を記憶することができる。言い換えれば、本発明のデバイスは、未処理のデータ（raw data）ではなく処理されたデータを記憶するので、より効率的にメモリを使用する。

20

【0014】

[0032]ローカルデバイスは、視聴者エンゲージメントを評価するために視聴エリアから取得された、視覚情報と深度情報の両方を含み得る、未処理の画像データを処理する。ローカルデバイスは、視聴者のボディジェスチャ、動き、および顔の向き（facial orientation）を分析するために、人工知能（AI）技術および機械学習技法を使用することができる。ローカルデバイスはまた、画像データから、ビデオオーディオの個々の顔を認識し、各視聴者の感情を決定することができる。この処理では、個人の画像は、個人の構内の外に送信されない。認識は、構内のローカルデバイス上で実行され得る。世帯中の各個人は、その世帯のためのオンボーディングプロセス中に一意の識別子を付与され得る。認識プロセス中に一致があったとき、この識別子が一致に割り当てられ、次いで、この識別子がリモートサーバに送信され得る。さらに、処理は、（画像を含む）ストリーミングビデオまたはオーディオデータ上で行われる。言い換えれば、ビデオまたはオーディオデータは、ローカルメモリに残存されない。

30

【0015】

[0033]ローカルデバイスは、視聴されている特定のビデオ（たとえば、TVチャンネル、番組、または広告）を識別するために、未処理のオーディオデータをオーディオデータベース中のサンプルと照合または比較することによって、未処理のオーディオデータを処理する。代替または追加として、ローカルデバイスは、オーディオが属するコンテンツの識別情報を識別して返すサードパーティアプリケーションプログラミングインターフェース（API）に、オーディオデータに基づくクエリをサブミットすることができる。いくつかの場合には、データベースまたはAPIは、複数的一致候補を返すことがあり、リモートサーバは、TVスケジュールに関する情報、後続のオーディオサンプル、あるいは限定はしないが、セットトップボックス、ケーブル/インターネット接続、またはコンテンツプロバイダ自体を含む他のソースから収集されたデータを使用して、最良の一致を選択することができる。

40

【0016】

[0034]いくつかの実装形態では、ローカルデバイスは、後の検索のために未処理の画像

50

またはオーディオデータを記憶しない。代わりに、ローカルデバイスは、処理のために未処理の画像およびオーディオデータを記憶する1つまたは複数のバッファに、未処理の画像およびオーディオデータを書き込み、次いで、未処理の画像およびオーディオデータが処理された後にバッファを上書きまたは消去する。言い換えれば、ローカルデバイスは、未処理の画像およびオーディオデータを、処理中に一時的に保持するにすぎない。本明細書で使用される、ローカルデバイスに未処理の画像およびオーディオデータを「保持する」ことは、(たとえば、それらの間の任意の値およびサブレンジを含む、100ミリ秒未満、80ミリ秒未満、60ミリ秒未満、50ミリ秒未満、または40ミリ秒未満の)短い持続時間の間の、これらのデータの一時的な記憶を指す。未処理の画像およびオーディオデータを上書きまたは消去することは、ローカルデバイスによって必要とされるメモリの量を低減することを含む、いくつかの利点を与える。それはまた、視聴エリア中またはマイクロフォンの範囲内の、子供を含む人々を識別するために使用され得る画像またはオーディオデータをなくすことによって、データプライバシー法の遵守をより容易にする。

10

【0017】

[0035]画像およびオーディオデータをローカルで処理および記憶することは、別の技術的利点を与え、すなわち、視聴習慣に関する情報をローカルデバイスからリモートサーバに伝達するために必要とされる帯域幅を低減する。未処理の画像およびオーディオデータと比較して、処理された画像およびオーディオデータは、より少ないメモリを消費し、したがって、送信のためにより小さい帯域幅を必要とする。処理された画像およびオーディオデータはまた、未処理の画像およびオーディオデータよりも緩やかに所与のメモリを埋め、したがって、より低い頻度でリモートサーバに送信され得る。ローカルデバイスは、ネットワーク帯域幅使用が比較的低い時間、たとえば、深夜または早朝にバースト送信をスケジュールすることによって、このフレキシビリティを利用し得る。視聴エリア中またはマイクロフォンの範囲内の、子供を含む人々を識別する情報を必ずしも含むとは限らない、処理された画像およびオーディオデータを送信することはまた、データプライバシー法の遵守の容易さを保証または向上させる。

20

【0018】

[0036]リモートサーバは、異なる世帯におけるローカルデバイスから、処理された画像およびオーディオデータを収集する。それは、コミュニティ中の異なる世帯から収集された視聴者エンゲージメント情報を統計的に分析することによって、コミュニティ全体にわたる視聴者エンゲージメントを評価するために、このデータを処理する。たとえば、サーバは、各世帯から収集された極めて細かいデータから、検出された番組の全長に対する、視聴者エンゲージメントの比を定量化することができる。

30

【0019】

[0037]統計的分析はさらに、ビデオを見ている人々および/または世帯中の人々の人口統計学的情報(たとえば、年齢、性別、世帯収入、民族など)を考慮に入れることができる。すべてのこの情報に基づいて、サーバは、視聴者エンゲージメントを定量化するために、視聴可能性インデックスおよび注意インデックス(attention index)(どちらも以下で定義される)など、様々なインデックスを計算し得る。これらの視聴者エンゲージメントインデックスは、視聴者のボディジェスチャ、動き、および視聴者の顔の向きに関する情報、ならびにビデオ情報を含む、ローカルデバイスによって提供される任意およびすべての情報に基づき得る。これらの定量的インデックスは、特に、(i)誰がディスプレイを実際に見ているか、(ii)オーディエンスメンバーがどのくらいの頻度でディスプレイを見るか、および(iii)ディスプレイ上の番組および広告に対するオーディエンスの反応を示すことができる。

40

【0020】

[0038]定量的インデックスは、次いで、リモートサーバによって中央ストレージ(たとえば、クラウドベースのデータベース)に転送され得、ここで、限定はしないが、TV広告代理店およびTVネットワークを含むサードパーティは、インデックスおよび場合によっては他のデータにもアクセスすることができる。代替的に、センサによって収集された

50

未処理データ (raw data) は、クラウド上の中央ストレージに転送され得、そこで、本明細書で説明される方法によって分析され、関係するサードパーティにとって利用可能になる。サードパーティは、オプションによりシステムを通して未処理データにアクセスし得る。この例における未処理データは、(ビデオおよびオーディオストリーム自体ではなく) ビデオおよびオーディオストリームを処理した後に収集されたデータを含む。概して、未処理データは、1秒未満ごとの(たとえば、1/2秒以下ごとの)、視聴者の一意の識別子、視聴者の注意深さ (attentiveness)、および視聴者によって視聴されている番組を含むことができる。さらなる定量的インデックス(以下のさらなる詳細を参照)が、この未処理データを使用してリモートサーバ上で算出され得る。

【0021】

[0039] 広告主にとって有益であり得る、個々の層 (demographics) の先例のない測定値を含む、この取得および分析されたデータは、コンテンツプロバイダまたは広告代理店などの収集エンティティが、ビデオの影響を正確に評価することを可能にすることができる。たとえば、広告代理店は、どのコマーシャル枠が広告代理店のターゲットオーディエンスに対する最良適合であるかを決定するために、データを使用することができる。人口統計学的情報を用いて、データは、オーディエンスのタイプと照合され得、効果的に購買行動を導き、それにより、番組における投資収益率 (ROI: return on investment) を高めることができる。TVネットワークも、そのTV番組のより正確なレーティング、オーディエンスタイプ、反応、および予測広告枠価値 (predictive ad slot value) を収集することができるので、データから恩恵を受けることができる。これはさらに、どの広告枠が特定のターゲット層に対して最も高い価値を有し得るかを決定することに加えて、TVネットワークがその番組をオーディエンスのタイプにより良く適合するように改善し、あまり人気がないショーをなくすことを可能にする。

【0022】

[0040] 取得および分析されたデータはまた、様々なビジネスモデルを可能にする。たとえば、収集エンティティは、国および/または地域の層を表す選択されたユーザ世帯に置かれた動き検知デバイスから収集された、分析のためのパフォーマンスベースのTVレーティングデータおよび未処理データを、TVネットワーク、広告代理店、および他の関係するサードパーティに、ならびに広告代理店からデータを取得する広告主に間接的に、提供することができる。

【0023】

[0041] 視聴者エンゲージメントを評価するシステム

[0042] 図1は、ディスプレイをもつ世帯、スポーツバー、または他の空間における視聴者エンゲージメントを評価するためのシステム100の概略図を示す。システム100は、視聴者エンゲージメントデータを収集するために各世帯に配設されたローカルデバイス105と、データを記憶するためのメモリおよびデータを分析するための(リモートプロセッサとも呼ばれる)プロセッサを含む、クラウドストレージおよびコンピューティングデバイスなど、リモートサーバ170を含む。ローカルデバイス105は、インターネット接続などのネットワーク接続172を介して、リモートサーバ170に通信可能に結合される。たとえば、ローカルデバイス105は、家庭用ローカルエリアネットワーク(LAN)に接続するためのWiFiアンテナまたはイーサネット(登録商標)ポートなど、ネットワークインターフェース165を含み得る。このLANは、たとえば、インターネットサービスプロバイダ(ISP)によって提供されるケーブルまたは光ファイバー接続を介して、ワイドエリアネットワーク(WAN)に接続される。

【0024】

[0043] 図1中のローカルデバイス105は、テレビジョン(TV)、コンピュータスクリーン、タブレット、または他のデバイスなど、ディスプレイ11の前の視聴エリア101を赤外線(IR)光で照射するためのIRエミッタ110を含む。このIR光は、視聴エリア101中の(人間のオーディエンスを含む)物体から散乱または反射する照射パターンを生成するために、構造化または調整され得る。ローカルデバイス105はまた、こ

10

20

30

40

50

これらの物体によって反射または散乱されたIR光を検出するIRセンサ120を含む。IRエミッタ110およびIRセンサ120に結合された(ローカルプロセッサ150とも呼ばれる)プロセッサ150は、視聴エリア101の1つまたは複数のIR深度画像またはIR深度マップを生成するために、照射パターンおよび検出されたIR光に関する情報を使用する。より具体的には、プロセッサ150は、反射されたビームから導出された情報を深度情報に変換し、視聴者とセンサ120との間の距離を測定する。プロセッサ150は、何人の人々が視聴エリアにいるか、およびそれらの人々のうちの誰がディスプレイを見ているかを決定するために、これらのIR深度画像を使用する。プロセッサ150はまた、ディスプレイを見ている人々の識別情報(identities)に関する情報を、IR深度画像から、場合によっては、その人々の顔またはジェスチャを認識することによって、あるいはその人々の層(たとえば、年齢、性別など)を決定することによって、導出し得る。

10

【0025】

[0044]ローカルデバイス105は、視聴エリア101のカラー画像をキャプチャする(可視カメラとも呼ばれる)RGBセンサ130をさらに含む。また、プロセッサ150は、RGBセンサに結合され、視聴エリア中にいる人々の数と、ディスプレイにエンゲージしている人々の数と、視聴エリア中の人々に関する情報とを推定するために、単独で、またはIR深度画像と組み合わせて、カラー画像を使用し得る。カラー画像は顔認識のためにも使用され得る。いくつかの場合には、プロセッサ150は、視聴エリア中の人々の数およびビデオにエンゲージしている人々の数の推定値の正確さを向上させるために、カラー画像とIR深度画像の両方を使用する。

20

【0026】

[0045]ローカルデバイス105はまた、ディスプレイ11に結合されたスピーカ13によって発せられた音を検出するように設置された1つまたは複数のマイクロフォン140を含む。動作中、スピーカ13は、ディスプレイ11上に示されているビデオのサウンドトラックを再生する。また、マイクロフォン140は、スピーカ13によって再生されたサウンドトラックのオーディオサンプルをキャプチャする。マイクロフォン140に結合されたプロセッサ150は、ビデオ(サウンドトラック)のオーディオフィンガープリントを作成するために、これらのオーディオサンプルを使用し、これを、ディスプレイ11上に示されているビデオを識別するために、所有者のまたはサードパーティのデータベース中の他のオーディオフィンガープリントと比較する。

30

【0027】

[0046]システム100は、Bluetooth(登録商標)送信機185に対応したBluetooth受信機180をさらに含むことができる。いくつかの場合には、Bluetooth送信機185は、視聴者によって装着されたリストバンドまたは腕時計に含まれ得る。動作中、Bluetooth送信機185は、Bluetooth受信機180によって受信される低電力Bluetoothビーコンを送信する。プロセッサ150は、次いで、受信されたBluetoothビーコンに基づいて、ディスプレイ11からの視聴者の距離を測定することができる。さらに、各Bluetooth送信機185は、プロセッサ150によって認識され得る一意のIDを有することができる。送信機IDはさらに、一意の視聴者に関連付けられ得る(たとえば、世帯中の各視聴者は、その人自身の送信機を有する)。このようにして、視聴者の識別情報も決定され得る。

40

【0028】

[0047]いくつかの場合には、システム100は、2つ以上のBluetooth受信機を含むことがある。これらの受信機は、各受信機が送信機185から異なるBluetooth信号強度を受信できるように、異なるロケーションに配設され得る。この構成は、プロセッサ150が、ディスプレイ11からの視聴者の距離だけでなく、(たとえば、ディスプレイ11の左側または右側の)視聴者の相対ロケーションをも推定することを可能にすることができる。

【0029】

[0048]システム100は、位置および動きを検出するために、3軸加速度計などの他の

50

動き検知デバイスを含み得る。動き検知デバイスは、たとえば、USBケーブルを介して、デスクトップマシンなど、データ分析および処理デバイスと接続され得る。

【0030】

[0049]図1は、ローカルデバイス105の一部として(たとえば、同じハウジング内)のデータ収集構成要素、ここでは、IREミッタ110、IRセンサ120、RGBセンサ130、およびマイクロフォン140を示す。他の実施形態では、これらの構成要素のうちの一つまたは複数は、USB接続、RS232接続、イーサネット接続、ファイバー接続など、一つまたは複数のワイヤード接続、あるいはWiFi接続、Bluetooth接続、他のRF接続、または赤外線接続など、一つまたは複数のワイヤレス接続によってプロセッサ150に結合された別個のデバイスとして実装され得る。たとえば、IREミッタ110およびIRセンサ120は、プロセッサ150に接続された、Microsoft(登録商標) Kinectなどの市販のデバイス(の中のもの)であり得る。同様に、マイクロフォン140は、視聴エリアの周りにまたはスピーカ13の近くに設置されたマイクロフォンのアレイとして実装され得る。マイクロフォンアレイは、環境雑音から音声入力をより良く抽出することが可能であり得る。ローカルデバイス105は、同様に、他のセンサを含むかまたはそれらに結合され得る。

10

【0031】

[0050]システム100中のプロセッサ150は、IREミッタ110と、IRセンサ120と、RGBセンサ130と、マイクロフォン140とを含むセンサによって取得された未処理データを処理するために採用される。処理は、プロセッサ150に結合されたメモリ160に記憶されたプロセッサ実行可能命令の実行時に行われ得る。一例では、リモートサーバ170から命令をダウンロードすることによって、ユーザが命令をメモリ160に手作業で記憶することができる。別の例では、ローカルデバイス105は、リモートサーバ170からダウンロードするために利用可能な更新された命令があるかどうかを(ルーチンの)確認するように構成され得る。そうである場合、ローカルデバイス105は、ネットワーク接続172およびネットワークインターフェース165を介して、更新を自動的にダウンロードすることができる。また別の例では、リモートサーバ170は、更新または新しい命令のセットがダウンロードの準備ができているとき、ローカルデバイス105に通知を送るように構成され得る。通知を受信すると、ユーザは、更新をダウンロードおよび/またはインストールするかどうかを決定することができる。また別の例では、リモートサーバ170は、スマートフォンなどの別のユーザデバイスに更新通知を送るように構成され得る。通知を受信すると、ユーザは、更新をダウンロードおよび/またはインストールするかどうかを決定することができる。

20

30

【0032】

[0051]ローカルデバイス105中のメモリ160はまた、処理されたデータ(たとえば、視聴エリア中の人々の数の推定値、ディスプレイにエンゲージしている人々の数の推定値、およびビデオの識別情報、ならびに未処理の画像およびオーディオデータから導出された人口統計学的情報またはインデックス)を記憶する。メモリ160が処理されたデータを十分に蓄積すると、プロセッサ150は、アグリゲーション、さらなる処理、および報告のために、ネットワークインターフェース165およびネットワーク接続172を介して、処理されたデータをリモートサーバ170に送信する。ローカルメモリ160はまた、ローカル処理中に、画像およびオーディオデータを一時的に保持する。いくつかの場合には、この処理は1/4秒未満で完了される。

40

【0033】

[0052]ローカルデバイスを用いて画像およびオーディオデータを収集および処理すること

[0053]図2Aは、図1に示されているシステム100のようなシステムを用いて取得された画像およびオーディオデータを収集および処理するためのプロセス200を示す。上記で説明されたように、システムは、ディスプレイの前の視聴エリアの画像に対し、可視センサ、IRセンサ、またはその両方を含めることができる(202)。一例では、RG

50

B センサ 1 3 0 および I R センサ 1 2 0 は、互いに独立して動作し、すなわち、センサは非同期的に画像を取得する。別の例では、R G B センサ 1 3 0 および I R センサ 1 2 0 による画像取得は、実質的に同期される。R G B センサ 1 3 0 が可視画像を取得するたびに、たとえば、同時にまたは交互に、I R センサ 1 2 0 が I R 画像を取得する。

【 0 0 3 4 】

[0054]ローカルプロセッサ（たとえば、プロセッサ 1 5 0 ）は、視聴エリアの画像中の人々の数を検出し（ 2 0 4 ）、またそれらの人々のうちの誰がディスプレイにエンゲージしているかを決定する（ 2 0 6 ）。たとえば、ローカルプロセッサは、コンピュータビジョン / 画像処理の技術分野で知られている骨格検出技法、顔認識技法、および視線追跡技法を含む、以下で説明される技法を使用し得る。いくつかの場合には、ローカルプロセッサ 1 5 0 は、各視聴者が視聴エリア中に存在する持続時間と、各視聴者がディスプレイにエンゲージしている持続時間と、表示されているビデオの識別情報とに關係する追加のインデックスを決定する（ 2 0 8 ）ことができ、ビデオの識別情報は、以下で説明されるようにオーディオデータから導出され得る（ 2 2 2 ）。

10

【 0 0 3 5 】

[0055]ローカルプロセッサはさらに、人口統計学的レベルで、視聴エリア 1 0 1 中の検出された各人物を識別することができる（たとえば、25 ~ 30 歳の男性、12 ~ 15 歳の少女）（ 2 1 0 ）。ローカルプロセッサ 1 5 0 が、たとえば、ローカルメモリ 1 6 0 またはリモートサーバ 1 7 0 を介して、ローカルデバイス 1 0 5 が設置された世帯に関する情報へのアクセスを有する場合、ローカルプロセッサ 1 5 0 は、視聴エリア 1 0 1 中の検出された各人物のより確信的な人口統計学的情報推定値を提供するために、この人口統計学的情報を使用し得る。ローカルプロセッサは、視聴エリアにいる世帯中の特定の人々を識別することさえある。

20

【 0 0 3 6 】

[0056]ローカルプロセッサ 1 5 0 はまた、視聴エリア 1 0 1 中の検出された各人物の気分または感情を推定することができる（ 2 1 2 ）。プロセッサ 1 5 0 によって決定され得る感情としては、たとえば、嬉しい、悲しい、またはどちらでもないがあり得る。ディスプレイ 1 1 上のビデオを見ているときの視聴者の感情の分類は、ビデオに対する視聴者の反応を測定するために使用され、それにより、広告のターゲット配信を促進することができる。

30

【 0 0 3 7 】

[0057]各人物の気分または感情を推定するために、ローカルプロセッサ 1 5 0 は、R G B チャネルと I R チャネルの両方からリアルタイムで、（たとえば、視聴エリア 1 0 1 の画像から）視覚情報をキャプチャすることができる。視覚情報は、異なる気分または感情状態のシグネチャであり得るパターンおよび特徴を抽出するために、さらに処理され得る。両方のチャネルから抽出された特徴は、統合された特徴として融合され得る。分類器は、入力としてそのような特徴をとるようにトレーニングされ得る。次いで、感情 / 気分の推定が、毎回、いくつかのパターンに対する分類器の応答に基づいて行われ得る。

【 0 0 3 8 】

[0058]いくつかの場合には、気分または感情の推定は、以下の方法によって達成され得る。本方法は、特に、笑っている、および眉をひそめているなど、人々が様々な感情を示しているトレーニング画像を収集することを含む。各感情を表す特徴が、これらのトレーニング画像から（たとえば、プロセッサによって）抽出される。特徴および画像は、次いで、分類器が各特徴を対応する感情に相関させるようにトレーニングするために使用される。このようにして、分類器は、これらの特徴を様々な感情に割り当てることができる。本方法はまた、リアルタイムで視聴者感情を認識するために、ローカルデバイス上に分類器を展開することを含む。

40

【 0 0 3 9 】

[0059]システムが同期的に可視および I R 画像を収集する場合、可視および I R カメラは、人々を検出し（ 2 0 4 ）、エンゲージしている視聴者をカウントし（ 2 0 6 ）、人口

50

統計学的に視聴者を識別し(210)、気分を推定する(212)ために、プロセッサによって使用されるコンピュータビジョンモデルをトレーニングするための画像を収集することができる。トレーニングは、「グラウンドトゥールズ(ground truth)」を確立するために採用され得る。ほとんど並行してIRセンサとRGBセンサの両方から画像データを収集すると、人間が、各画像中の検出された人々に注釈を付けることができる。この手作業のデータは、トレーニングアルゴリズムに供給され得、2つの別個のモデル、可視RGBスペクトルに関してトレーニングされた一方と、IRスペクトルに関してトレーニングされた他方とをもたらす。次いで、「グラウンドトゥールズ」に対する各モデルの検出レートは、より良く機能するモデルを選択するために比較される。このトレーニングのさらなる詳細は、図2Bを参照しながら以下で説明される。

10

【0040】

[0060]また、2つのカメラ(たとえば、図1中のセンサ120および130)の同期は、ローカルプロセッサが画像処理をダブルチェックすることを可能にすることができる。たとえば、プロセッサ150は、各画像中の識別された人々の数を比較するか、あるいは一方の画像では見えるが他方の画像では見えにくいまたは見えないエラーを削除することができる。結果が互いに一致している場合、プロセッサ150は結果を記録することができる。そうでない場合、プロセッサ150は、画像のうちの少なくとも1つにおける、可能性のあるエラー(possible errors)を検出することができる。代替的に、プロセッサ150は、人間が介入するためのアラートを生成することができる。プロセッサ150はまた、これらの2つの画像から推定されたデータに関連付けられたフラグを生成し、このデータがあまり信頼できないことがあることを示すことができる。後続の分析では、問題となっている画像のこのペアの少し前または少し後に撮影された画像が、信頼できる人物認識を提供することができる場合、このデータはまったく使用されないことがある。

20

【0041】

[0061]一例では、ローカルデバイス105は、画像データを撮影するために、常に可視センサ120およびIRセンサ130を使用する。別の例では、ローカルデバイス105は、画像データを撮影するために、センサ120またはセンサ130のうちの1つのみを使用することができる。また別の例では、ローカルデバイス105は、一方のセンサをデフォルトセンサとして使用し、他方のセンサをバックアップセンサとして使用することができる。たとえば、ローカルデバイス105は、画像撮影のために、大部分の時間RGBセンサ130を使用することができる。しかしながら、プロセッサ150が可視画像を満足できる程度に分析するのが困難な(たとえば、分析が所望されるほど信頼できるものではない)場合、プロセッサ150は、バックアップとしてIRセンサ120をオンにすることができる(またはその逆も同様である)。これは、たとえば、視聴エリアにおける周囲光レベルが低いときに起こり得る。

30

【0042】

[0062]ローカルプロセッサはまた、視聴エリア中の人々の数と、視聴エリア中の人々の位置と、ディスプレイ上のビデオの識別情報とに基づいて、可視センサ、IRセンサ、またはその両方のための画像取得レートを調整し得る(214)。概して、いずれか一方または両方のセンサについての画像取得は、実質的に毎秒約15フレーム(fps)以上であり得る(たとえば、それらの間の任意の値およびサブレンジを含む、約15fps、約20fps、約30fps、約50fpsまたはさらに大きい)。この画像取得レートにおいて、センサは、ローカルプロセッサが視聴者エンゲージメントを評価するのに十分なだけ、眼球運動を検出することができる(206)。

40

【0043】

[0063]ローカルプロセッサは、視聴エリア101中の人々の数に基づいて、画像取得レートを増加または減少させ得る。たとえば、プロセッサが視聴エリア101に誰もいないと決定した場合、プロセッサは、電力およびメモリ消費を低減するために、画像取得レートを低減し得る。同様に、(たとえば、視聴者が眠っているように見えるので)視聴者がビデオにエンゲージしていないとプロセッサが決定した場合、プロセッサは、電力、メモ

50

り、またはその両方を節約するために、画像取得レートを低減し得る。逆に、視聴者がその注意 (attention) をすばやくシフトしているように見える場合、視聴者がテンポの速いビデオ (たとえば、フットボールの試合またはアクション映画) を見ている場合、視聴者がすばやくチャンネルを変更している (たとえば、チャンネルサーフィンをしている) 場合、または (たとえば、一連の広告の間に) コンテンツが比較的すばやく変化している場合、プロセッサは、画像取得レートを (たとえば、15 fps よりも大きく) 増加させ得る。

【0044】

[0064]システムがIR画像センサと可視画像センサの両方を含む場合、ローカルプロセッサはまた、照明条件または相対的画像品質に基づいて、画像取得を変化させ得る。たとえば、低光量条件では、ローカルプロセッサは、可視画像よりも高いレートでIR画像を取得し得る。同様に、ローカルプロセッサが、可視画像を処理してIR画像よりも良好な結果を得る場合、ローカルプロセッサは、IR画像よりも高いレートで可視画像を取得し得る (または反対が真の場合、その逆も同様である)。

10

【0045】

[0065]システムはまた、マイクロフォン140を用いて、ビデオのサウンドトラックのサンプルを記録する(220)。概して、オーディオデータ取得レートまたはオーディオサンプリングレートは、画像取得レートより低い。たとえば、マイクロフォンは、30秒ごとに1回のレートでオーディオサンプルを取得する。各取得では、マイクロフォン140は、オーディオサンプルに関連付けられたビデオの識別を可能にするように、有限持続時間を有するオーディオサンプルを記録する。オーディオサンプルの持続時間は、実質的に5秒以上であり得る (たとえば、それらの間の任意の値およびサブレンジを含む、約5秒、約6秒、約8秒、約10秒、約20秒、または約30秒)。

20

【0046】

[0066]ローカルプロセッサは、ディスプレイ上で再生されているビデオを識別するために、マイクロフォン140によって記録されたオーディオサンプルを使用する(222)。たとえば、プロセッサ150は、オーディオデータのフィンガープリントを作成し、サードパーティアプリケーションプログラミングインターフェース(API)にクエリを実行するために該フィンガープリントを使用することができ、サードパーティアプリケーションプログラミングインターフェース(API)は、そのクエリに対して、オーディオデータに関連付けられたビデオの識別情報で応答する。別の例では、プロセッサ150は、ビデオの識別情報を決定するために、フィンガープリントをローカルテーブルまたはメモリと比較することができる。

30

【0047】

[0067]上述のように、ビデオを識別するためにビデオサウンドトラックのサンプルを使用することは、ビデオを識別するための従来のTV調査デバイスによって使用されるデジタルウォーターマークに勝るいくつかの利点を与える。それは、デジタルウォーターマークをビデオ中に挿入することを必要とせず、コンテンツ製作者およびプロバイダと協調する必要をなくす。これにより、コンテンツ制作および配信が簡略化され、デジタルウォーターマークを提供することができないかまたは提供しない製作者および配信者を含む、より広範囲のビデオコンテンツを識別および評価することが可能になる。また、ローカルデバイスをケーブルまたはセットトップボックスに接続する必要をなくす。

40

【0048】

[0068]さらに、デジタルウォーターマークの代わりにオーディオデータを使用することは、「フォールスポジティブ」のリスク、またはシステムが視聴エリア中の人々を検出し、TVがオフのときでも実際に見られていないビデオを識別する場合の事例を低減する。これは、世帯のメンバが、TVがオフのときでもセットトップボックスをオンのままにした場合に、セットトップボックスに取り付けられた従来のシステムで起こることがある。

【0049】

[0069]いくつかの例では、ローカルプロセッサは、たとえば、ビデオの識別情報、視聴

50

エリア中の人々の数、ビデオにエンゲージしている人々の数などに基づいて、オーディオサンプリングレートを調整する(224)。たとえば、ローカルプロセッサが単一のフィンガープリントからビデオを識別することができない場合(たとえば、ビデオサウンドトラックが、多くの異なるビデオサウンドトラック中に現れる流行歌を含むという理由で)、ローカルプロセッサおよびマイクロフォンは、あいまいさを解決するようにビデオを改善するために、より高いレートで、またはより長い持続時間のサンプルを取得し得る。プロセッサはまた、視聴エリア101に誰もいないか、または(たとえば、視聴者が眠っているように見えるので)視聴者がビデオにエンゲージしていない場合に、電力、メモリ、またはその両方を節約するために、オーディオサンプリングレートを減少させ得る。逆に、プロセッサは、視聴者がすばやくチャンネルを変更している(たとえば、チャンネルサーフィンをしている)場合に、または(たとえば、一連の広告の間に)コンテンツが比較的すばやく変化している場合に、オーディオサンプリングレートを増加させ得る。

10

【0050】

[0070]実装形態に応じて、マイクロフォンは、一定の間隔で(すなわち、周期的に)、または不規則な間隔で(たとえば、非周期的にまたは時間変動する周期で)、オーディオサンプルを記録し得る。たとえば、マイクロフォンは、1日を通して一定のレート(たとえば、1分につき約2つのサンプル)で、オーディオデータを取得し得る。他の場合には、マイクロフォンは、TVがオンであるかオンになりそうなとき(たとえば、夕方)、あるサンプリングレートで動作し、TVがオフであるかオフになりそうなとき(たとえば、早朝、日中)、別の低いサンプリングレートで動作し得る。ローカルプロセッサが、オーディオサンプルからTVがオン(オフ)にされたことを検出した場合、ローカルプロセッサは、それに応じてサンプルレートを増加(減少)させ得る。また、オーディオサンプルからTVがオン(オフ)にされたことを検出したことに応答して、視聴エリアを撮像するのを開始(停止)するように画像センサをトリガし得る。

20

【0051】

[0071]未処理の画像およびオーディオデータが処理されると、ローカルプロセッサは、未処理の画像およびオーディオデータを上書きするか、または未処理の画像およびオーディオデータをメモリから消去する(230)。言い換えれば、プロセッサ150が人間を検出および識別し、そのエンゲージメントおよび表情を測定する間、各画像はメモリ150に保持される。検出、識別、およびエンゲージメントデータはフレームごとに収集され、この情報は残存され、最終的にバックエンドサーバ170にアップロードされる。同様に、サードパーティAPIがオーディオフィンガープリントを処理し、関連するビデオの識別情報を返す間、オーディオデータもメモリ160に保持される。識別情報は、以下で説明されるように、記憶されるか、および/またはバックエンドサーバ170にアップロードされる。

30

【0052】

[0072]未処理の画像およびオーディオデータを上書きまたは消去(または場合によっては廃棄)することによって、ローカルプロセッサは、メモリの負担を低減し、視聴エリア中の個人を識別する能力を低減するかまたはなくす。このことは、システムをハッキングする潜在的試みに対してより少ない情報をさらすことによって、個人のプライバシーを守る。それはまた、個人の画像をサードパーティに送信する可能性をなくす。これは、児童オンラインプライバシー保護法(Children's Online Privacy Protection Act)による、視聴エリア中の子供のプライバシーを守るために特に有益である。

40

【0053】

[0073]いくつかの場合には、ローカルプロセッサは、未処理の画像およびオーディオデータをメモリからアクティブに消去する。他の場合には、ローカルプロセッサは、あらかじめ決定された量(たとえば、1つの画像または1つのオーディオサンプル)を超える未処理の画像およびオーディオデータを記憶しないようにサイズ決定されたメモリ中の1つまたは複数のバッファに、その未処理の画像およびデータを記憶する。ローカルプロセッサは、次の画像またはオーディオサンプルがバッファを上書きするように、サンプル間の

50

時間期間中に未処理の画像およびデータを分析する。

【 0 0 5 4 】

[0074]ローカルプロセッサ 1 5 0 はまた、処理されたデータをメモリ 1 6 0 に記憶する。処理されたデータは、メモリ要件を低減するために、カンマ区切り変数 (C S V) などの比較的コンパクトなフォーマットで記憶され得る。C S V または他のファイル中に含まれるデータは、たとえば、各画像中に誰かが存在するかどうか、各画像中の視聴エリア 1 0 1 中の人々の数、視聴エリア 1 0 1 中のディスプレイ 1 1 を実際に見ている人々の数、各視聴者の感情の分類、および各視聴者の識別情報を示し得る。処理されたデータはまた、I R 画像取得レート、可視画像取得レート、オーディオサンプリングレート、現在のソフトウェア / ファームウェアアップデートなどを含む、ローカルデバイスの動作状態に関する指示を含み得る。

10

【 0 0 5 5 】

[0075]ローカルプロセッサは、記憶のためにまたはさらなる処理のために、処理されたデータを (たとえば、ネットワークインターフェースを介して) リモートサーバに送信する (2 3 6)。処理されたデータが比較的コンパクトなフォーマットであるので、アップロード帯域幅は、未処理の画像およびオーディオデータの場合よりも、はるかに低い。また、送信されたデータは、視聴エリアの画像、または視聴者の音声を含み得るオーディオサンプルを含まないので、視聴者のプライバシーを脅かすリスクがあまりない。さらに、未処理の画像およびオーディオ画像がリモートサーバに送信され、該リモートサーバによって処理される場合よりも、処理されたデータのオーディオおよび画像部分は、それらがローカルに処理されるので、同期されており、引き続きそのままである可能性が高い。

20

【 0 0 5 6 】

[0076]いくつかの場合には、ローカルプロセッサは、処理されたデータを、それが処理されるにつれてリモートに送信し得る。他の場合には、ローカルプロセッサは、たとえば、利用可能なアップストリーム帯域幅、データの量などに基づいて、送信ウィンドウを識別し得る (2 3 4)。これらの送信ウィンドウは、あらかじめ決定されるか (たとえば、E T 午前 2 時)、ローカルデバイスの取付け中に世帯のメンバによって設定されるか、(たとえば、ソフトウェアまたはファームウェアアップデートを介して) リモートサーバによって設定されるか、または帯域幅測定値に基づいてローカルプロセッサによって決定され得る。

30

【 0 0 5 7 】

[0077]図 2 B は、視聴者エンゲージメントを定量化するためのコンピュータビジョンモデルをトレーニングする方法を示す。2 4 1 において、R G B センサと I R センサの両方は、2 つのタイプの処理を受けるビデオデータを取得する。2 4 2 a において、ビデオデータは、各フレーム中の顔を識別するために、手作業で注釈を付けられる。2 4 2 b において、現在のモデル (たとえば、デフォルトモデルまたは前の使用からのモデル) が、各フレーム中の顔を自動的に検出するために s u e d される。2 4 3 b において、2 4 2 a において取得された注釈付きビデオに対する、2 4 2 b における自動検出の精度を算出するために、プロセッサが使用される。2 4 4 において、精度が許容できる場合、方法 2 4 0 は 2 4 5 に進み、ここで、現在のモデルは (たとえば、方法 2 0 0 において使用される) 顔認識のためのプロダクションモデルとして設定される。精度が許容できない場合、方法 2 0 0 は 2 4 3 a に進み、ここで、ビデオは、ビデオのトレーニングセット (2 4 6 a) とビデオのテストセット (2 4 6 b) とに分割される。たとえば、R G B ビデオがトレーニングビデオ 2 4 6 a として選択され得、I R ビデオがテストビデオ 2 4 6 b として選択され得る (またはその逆も同様である)。

40

【 0 0 5 8 】

[0078]トレーニングビデオ 2 4 6 a は、2 4 7 a において新しいモデルをトレーニングするために送られ、一方、テストビデオ (2 4 6 b) は、新しいモデルをテストするためにステップ 2 4 7 b に送られる。2 4 7 b において、トレーニングビデオ 2 4 6 a およびテストビデオ 2 4 6 b は、2 4 7 c において新しいモデルの精度を算出するために、一緒

50

に収集される。249において、プロセッサは新しいモデルの精度を再び算出する。精度が許容できる場合、新しいモデルはプロダクションモデルとして設定される(245)。そうでない場合、方法240は248に進み、ここで、新しいモデルのパラメータが調整される。代替的に、248において、別の新しいモデルが構築され得る。いずれの場合も、新しいモデルのパラメータが247aに送られ、ここで、トレーニングビデオ246aが、新しいモデルをトレーニングするために使用される。このようにして、許容できる精度を有するように、新しいモデルが反復的に構築され得る。

【0059】

[0079]リモートサーバ動作

[0080]動作中、リモートサーバ170は、異なる世帯に配設された異なるローカルデバイス105から送信されたデータを収集する。リモートサーバ170は、定期的に着信データを読み取ることができる。リモートサーバ170はまた、受信データを構文解析し、各々が保存されたときのタイムスタンプを使用して、ビデオ認識データをオーディオ認識データと結びつけることができる。

10

【0060】

[0081]リモートサーバ170はまた、誤ったラベルを付けられたデータを正すことができる。たとえば、リモートサーバ170は、視聴者が識別されないかまたは誤識別されたときに、先行するまたは後続のタイムスタンプからのデータを使用してブリップを修正する(fix blips)ことができる。人物が、問題となっている画像に先行する画像中で識別され、問題となっている画像に後続する画像中でも識別された場合、リモートサーバ170は、この人物が問題となっている画像にも現れると決定することができる。

20

【0061】

[0082]リモートサーバ170はまた、ローカルデバイス105から受信されたデータおよび/またはリモートサーバ170によって処理されたデータを、照会可能データベース(query-able database)にロードすることができる。一例では、リモートサーバ170はまた、ユーザにアクセスを与えることができ、その場合、ユーザは記憶されたデータを分析のために使用することができる。別の例では、照会可能データベース中の記憶されたデータはまた、リモートサーバ170によって実行されるさらなる分析を可能にすることができる。たとえば、リモートサーバ170は、データベースを使用して、注意インデックスおよび視聴者インデックスを計算することができる。

30

【0062】

[0083]視聴者エンゲージメントを評価すること

[0084]図3A~図6は、視聴可能性インデックスおよび注意インデックスなどの尺度を使用して、ビデオとの視聴者エンゲージメントを定量化する方法を示す。以下の定義は、ビデオとの視聴者エンゲージメントを定量化するための本発明の方法および装置を理解するのに役立つ。

【0063】

[0085]番組持続時間(Program Duration)は、たとえば、秒、分、または時間単位の、一意の番組の合計持続時間として定義される。異なる番組の持続時間が比較され得る限り、使用される実際の単位(秒、分、または時間)は重要でない。

40

【0064】

[0086]コマーシャル持続時間(Commercial Duration)は、一意のコマーシャルの(たとえば、秒または分単位の)合計持続時間として定義される。

【0065】

[0087]ウォッチング持続時間(Watching Duration)(秒)は、世帯ごとの、一意の番組またはコマーシャルの見られた合計持続時間(秒数)として定義される。代替的に、ウォッチング秒(Watching Second)は、秒単位の番組の合計持続時間から、どの世帯も番組を見ていない間の(秒単位の)合計時間を引いたものとして定義され得る。

【0066】

[0088]総合ウォッチング持続時間(Aggregated Watching Duration)(秒)は、す

50

べての世帯にわたる、一意の番組またはコマーシャルの見られた合計持続時間（秒数）として定義される。

【0067】

[0089]ポジティブ持続時間比（Positive Duration Ratio）は、見られた番組またはコマーシャル広告のパーセンテージ（％）として定義される。より具体的には、番組または広告のポジティブ持続時間比は、番組または広告の合計持続時間×世帯の数分の、総合ウォッチング持続時間の比として計算され得る。

【0068】

[0090]視聴者カウント（VC：Viewer Count）は、所与の番組またはコマーシャル広告について正のウォッチング秒をもつすべての世帯にわたる、視聴エリア中の視聴者の合計数として定義される。

10

【0069】

[0091]ウォッチングレート（WR：Watching Rate）は、すべての世帯中の人々の合計数分の、TVがオンであるすべての世帯にわたる人々の合計数の比として定義される。たとえば、本方法が合計300人を有する100世帯を考慮に入れる場合。100人を有する30世帯が、TV受像機をオンにした場合、ウォッチングレートは、33.3％（すなわち、100/300）である。ただし、同じ30世帯が150人を有する場合、ウォッチングレートは50％（すなわち、150/300）である。

【0070】

[0092]視聴レート（VR：Viewing Rate）は、オンであるTV受像機の合計数分の、すべての世帯にわたる視聴エリア中の人々の合計数の比として定義される。たとえば、40個の異なるTV受像機によって定義される視聴エリアに100人がいる場合（各TV受像機は1つの視聴エリアを定義する）、視聴レートは2.5（すなわち、100/40）である。

20

【0071】

[0093]注意レート（AR：Attention Rate）は、すべての世帯にわたる視聴エリア中の人々の合計数分の、すべての世帯にわたるTVに注意している人々の合計数の比として定義される。たとえば、本方法によって考慮に入れられるすべての個人にわたって100人が視聴エリア中にいるが、60人のみがTVを実際に見ている（残りの40人は、他のことをしながらただTVをオンにまかにし得る）場合、注意レートは0.6または60％である。

30

【0072】

[0094]視聴可能性インデックス（VI：Viewability Index）は、各番組およびコマーシャルについての視聴レート（VR）の平均として定義される。

【0073】

[0095]注意インデックス（Attention Index）は、各番組およびコマーシャルについての注意レート（AR）の平均として定義される。

【0074】

[0096]図3Aは、顔および眼球追跡310、顔認識320、ならびに感情分析330を含む、視聴者エンゲージメントを評価する方法300（たとえば、図2Aの方法200におけるボックス206）を示す。プロセッサ（たとえば、図1に示されているローカルプロセッサ150）が、方法300を実装するために使用され得る。方法300における入力データは、視聴エリアの画像データ、オーディオデータ、または深度データなど、図1に示されているローカルデバイス105によって取得されたデータであり得る。顔が動くにつれてそれを追跡するための特徴的データポイントを識別し、ユーザがスクリーンを見ているかどうかを決定するために、顔および眼球追跡310が採用される。たとえば、人工知能を使用して、視聴者の識別情報を決定するために、顔認識320が採用される。たとえば、特に、顔の特徴、ジェスチャ、および心拍数を分析するための人工知能を使用して、視聴者の感情を決定するために、感情分析330が採用される。

40

【0075】

50

[0097]視聴者がスクリーンを実際に見ているかどうか、視聴者の識別情報、および視聴者の感情を含む、取得された情報は、様々なビデオレーティング340を決定するために使用される。一例では、取得された情報は、各世帯について個々のビデオレーティングを推定するために使用される。別の例では、取得された情報は、各人口統計学的領域について個々のビデオレーティングを推定するために使用される。また別の例では、取得された情報は、ビデオのグループについて全体的ビデオレーティングを推定するために使用される。また別の例では、取得された情報は、特定のビデオ（たとえば、番組および広告）に対するオーディエンス反応を推定するために使用される。取得された情報はまた、以下で説明されるように、視聴可能性インデックスおよび注意インデックスなど、視聴者エンゲージメントの定量的尺度を決定するために使用され得る。

10

【0076】

[0098]方法300におけるステップ310、320、および330は、パターン認識技法を使用して達成され得る。これらの技法は、たとえば、1つまたは複数の人間の顔を認識することによって、視聴者が視聴エリア中に存在するかどうかを決定することができる。実際に顔が認識された場合、これらの技法は、たとえば、認識された顔を、ビデオが再生している世帯の顔データを含むデータベースと比較することによって、視聴者が誰であるかをさらに決定することができる。代替的に、これらの技法は、視聴者が世帯からではない場合に備えて、より多くの人々（たとえば、可能な場合コミュニティ全体）の顔データを含むように拡張されたデータベースを使用し得る。これらの技法はまた、たとえば、視聴者がビデオを見ているかどうかを決定するために、顔の動きを追跡し、顔の向きを分析することができる。

20

【0077】

[0099]人工知能、機械学習、およびトレーニングされたニューラルネットワーク学習技法も、視聴者の感情を分析するために使用され得る。この目的で、これらの技法は、特に、ボディジェスチャ（ある時点における静的ジェスチャ）、身体の動き（ジェスチャの変化）、顔の向き、顔の方向（direction）/動き/位置、および心拍数を分析する。

【0078】

[0100]別の例では、方法300は、たとえば、図1に示されているRGBセンサ140およびIRセンサ120によって取得された画像データから、最初に顔を認識することができる。方法200はまた、顔の位置を検出し、顔の特徴的ポイント（たとえば、図2Aに示されている眼および口の境界ポイント）を識別し、顔が動くにつれてそれを追跡することができる。眼球追跡技法を使用して、方法300は、ビューがビデオを実際に見ている（または、代わりに、ただ視聴エリアに座っているが、何か他のことをしている）かどうか、を決定することができる。次いで、トレーニングされたニューラルネットワーク学習の技法を使用して、方法300は、同様の位置にあるデータベースからの顔の特徴を比較することによって、視聴者を世帯中の知られている人物と照合することができる。視聴者が識別されると、方法300は、ユーザの気分および/または感情を決定するために、顕著な顔の構成について視聴者を継続的に追跡することができる。

30

【0079】

[0101]方法300はまた、特定のタイミングポイントにおいてどのビデオが再生されているかを確かめるために、（たとえば、図1に示されているマイクロフォン140によって取得された）オーディオデータを、ビデオ（たとえば、TVショー）および他のオーディオのオーディオデータベースと比較することができる。一例では、ビデオ照合（video matching）は、方法300によって識別された視聴者によってどのTV局が視聴されているかを決定することができる。別の例では、ビデオ照合は、視聴者によってどのTV番組が視聴されているかを決定することができる。また別の例では、ビデオ照合は、どのコマーシャル広告が視聴されているかを決定することができる。代替または追加として、視聴されているTVチャンネル、番組、または広告は、限定はしないが、ケーブルまたは衛星放送用セットトップボックス、あるいは他の番組プロバイダのハードウェアまたはブロードキャスト信号を含む、他のソースから収集されたデータから決定され得る。

40

50

【 0 0 8 0 】

[0102]図 3 B は、本明細書で説明される視聴者エンゲージメントを定量化するための技法を介して推定され得る、視聴可能性インデックスおよび注意インデックスの概念を示す。概して、視聴可能性インデックスは、スクリーン上にあるものが人々を部屋に連れてくる傾向を定量化する。注意インデックスは、スクリーン上にあるものが視聴オーディエンスをエンゲージさせる傾向を定量化する。言い換えれば、視聴可能性インデックスは、ビデオ（または他の表示されたコンテンツ）が最初に視聴者を引きつける確率と考えられ得、注意インデックスは、視聴者がすでに視聴エリアにいるときから、ビデオが視聴者をディスプレイの前にとどめる確率と考えられ得る。図 3 B に示されているように、視聴可能性インデックスは、視聴エリア中に存在する人々の数に依存し、注意インデックスは、ディスプレイを実際に見ている人々数に依存する。

10

【 0 0 8 1 】

[0103]視聴可能性インデックスおよび注意インデックスを用いて視聴者エンゲージメントを評価すること

[0104]図 4 A は、視聴可能性インデックスを使用して視聴者エンゲージメントを定量化する方法 4 0 1 を示す。方法 4 0 1 はプロセッサによって実装され得る。方法 4 0 1 は、たとえば、図 1 に示されているシステムにおけるローカルデバイス 1 0 5 を取り付けるかまたは使用することを介して本方法に参加する複数の世帯中の各世帯において、プロセッサによって画像データが取得される、ステップ 4 1 1 において開始する。画像データは、ビデオ（たとえば、TV 番組、広告、ユーザリクエストビデオ、または他の任意のビデオ）を再生することができるディスプレイの前の視聴エリアの画像を含む。さらに、プロセッサはまた、ステップ 4 1 1 において、ディスプレイがビデオを示しているかどうかを決定する。ステップ 4 2 1 において、プロセッサは、ディスプレイによって再生される各ビデオについて、視聴レートおよびウォッチングレートを推定する。視聴レートは、上記で定義したように、ビデオを示しているディスプレイの合計数に対する、視聴エリア中の人々の合計数の比を表す。同様に、ウォッチングレートは、上記で定義したように、複数の世帯中の人々の合計数に対する、ディスプレイがビデオを示している世帯中の人々の合計数の比を表す。

20

【 0 0 8 2 】

[0105]視聴レートおよびウォッチングレートの推定は、ステップ 4 1 1 において取得された画像データと、複数の世帯中の各世帯に関する人口統計学的情報とに基づく。人口統計学的情報は、プロセッサが人口統計学的情報を容易に検索することができるように、プロセッサに動作可能に結合されたメモリに記憶され得る。別の例では、プロセッサは、別のサーバから人口統計学的情報を取得することができる。ステップ 3 3 0 において、プロセッサは、複数のビデオ中の各一意のビデオについて、視聴レートおよびウォッチングレートに基づいて視聴可能性インデックスを決定する。視聴可能性インデックスは、番組およびコマーシャルなど、各ビデオについての視聴レートの平均として上記で定義される。

30

【 0 0 8 3 】

[0106]方法 4 0 1 は、ディスプレイによって再生された各ビデオの視聴者カウントおよびポジティブ持続時間比を推定することをさらに含むことができる。該推定は、画像データと、複数の世帯中の各世帯に関する人口統計学的情報とに基づく。上記で定義したように、視聴者カウントは、各一意のビデオにエンゲージしている人々の合計数を表し、ポジティブ持続時間比は、一意のビデオの持続時間に対する、複数の世帯中の人々が一意のビデオを見ることによって費やされた合計時間の比を表す。

40

【 0 0 8 4 】

[0107]視聴者カウントおよびポジション持続時間比に基づいて、均衡視聴可能性インデックス (balanced viewability index) が決定され得る。一例では、均衡視聴可能性インデックスは、所与の各番組およびコマーシャルについて、視聴者カウントおよびポジティブ持続時間比を考慮に入れることによる、視聴可能性インデックス (V I) の加重平均として計算され得る。別の例では、均衡視聴可能性インデックスは、複数のビデオ中の一

50

意のビデオにわたる視聴可能性インデックスを正規化することによって、計算され得る。

【0085】

[0108]方法401は、平均視聴可能性インデックスを生成するために、有限時間期間の間のすべての番組およびコマーシャルにわたる視聴可能性インデックスを平均化することをさらに含むことができる。各番組およびコマーシャルの視聴可能性インデックスは、広告代理店、TV局、または他のコンテンツプロバイダなど、ユーザのための最終的な視聴可能性インデックス（無次元量）を生成するために、（たとえば、毎日、毎週、または毎月、算出される）平均視聴可能性インデックスで除算され得る。一例では、有限時間期間は、約2週間である。別の例では、有限時間期間は、約1か月である。また別の例では、有限時間期間は、約3か月である。

10

【0086】

[0109]画像データは、様々な取得レートで取得され得る。一例では、画像データは毎秒50回撮影され得る（50Hz）。一例では、画像データは毎秒30回撮影され得る（30Hz）。また別の例では、画像データは1秒ごとに撮影され得る（1Hz）。また別の例では、画像データは2秒ごとに撮影され得る（0.5Hz）。また別の例では、画像データは5秒ごとに撮影され得る（0.2Hz）。さらに、方法300は、世帯の人口統計学的情報を考慮に入れて視聴者エンゲージメント情報を導出するように、視聴エリア中の視聴者ごとに画像データを撮影および分類することができる。

【0087】

[0110]図4Bは、注意インデックスを使用してビデオとのユーザエンゲージメントを定量化する方法402を示す。方法402は、視聴者エンゲージメント評価に参加する各世帯について、ディスプレイの前の視聴エリアの画像データが撮影される、ステップ412を含む。ステップ412において、プロセッサは、（たとえば、図1に示されているローカルデバイス105中のマイクロフォン140によって取得されたオーディオデータを介して）画像データが撮影されたとき、ディスプレイがビデオを示しているかどうかを決定する。ステップ422において、ディスプレイによって再生された各ビデオについて、プロセッサは、画像データと世帯に関する人口統計学的情報とに基づいて注意レートを推定する。上記で定義したように、注意レートは、視聴エリア中の人々の合計数に対する、ビデオにエンゲージしている人々の合計数の比を表す。ビデオの注意レートに基づいて、ステップ432において、ビデオの有効性を示すために注意インデックスが決定される。

20

30

【0088】

[0111]方法402は、ディスプレイによって再生されたビデオの視聴者カウントおよびポジティブ持続時間比（positive duration ratio）を推定することをさらに含むことができる。方法401と同様に、方法402は、画像データと各世帯に関する人口統計学的情報とに基づいて、視聴者カウントおよびポジティブ持続時間レシオン（positive duration ration）を決定することができる。視聴者カウントおよびポジティブ持続時間レシオンを使用して、プロセッサは、次いで、均衡注意インデックスを決定することができる。方法402は、所与の時間期間（たとえば、1週間または1か月）にわたって、複数のビデオ中の一意のビデオにわたる注意インデックスを正規化することによって、正規化された注意インデックスを生成することを含むことができる。

40

【0089】

[0112]方法402は、平均注意インデックスを生成するために、有限時間期間の間のすべての番組およびコマーシャルにわたる注意インデックスを平均化することをさらに含むことができる。各番組およびコマーシャルの注意インデックスは、広告代理店、TV局、または他のコンテンツプロバイダなど、顧客のための最終的な注意インデックス（無次元量）を生成するために、平均注意インデックスで除算され得る。

【0090】

[0113]顔認識技法を使用して視聴者エンゲージメントを評価すること

[0114]図5は、顔認識技法および他の人工知能技法を使用して、ビデオとの視聴者エンゲージメントを評価する方法を示す。方法500は、（たとえば、図1に示されているシ

50

ステムを使用して)ディスプレイの前の視聴エリアの画像がキャプチャされる、ステップ510において開始する。取得された各画像について、視聴エリア中の人々の数がステップ520において推定される。一例では、推定は、たとえば、顔認識技法を使用して実行され得る。別の例では、推定は、身体骨格検出に基づいて実行され得る。

【0091】

[0115]ステップ530において、ディスプレイに対して、視聴エリア中の各人物の顔の向きが決定される。たとえば、顔の向きはディスプレイに向かっていることがあり、これは、視聴者がディスプレイ上のビデオを実際に見ていることを示し得る。代替的に、顔の向きはディスプレイから離れていることがあり、これは、視聴者がディスプレイの視聴エリア内にいるが、その人はビデオを見ていないことを示し得る。したがって、視聴者の顔の向きに基づいて、プロセッサは、ステップ540において、視聴エリア中の各人物が実際にビデオにエンゲージしているかどうかを評価することができる。ビデオを実際に見ている人々を見ていない人々と区別することによって、プロセッサは、ビデオの有効性のより正確な決定を行うことができる。ビデオの有効性は、たとえば、ビデオがどのくらいの時間の間、視聴者をエンゲージしている状態に保つことができるかによって、定量化され得る。

10

【0092】

[0116]骨格、顔、識別情報、感情、およびエンゲージメントを検出すること

[0117]図6は、骨格、顔、識別情報、感情、およびエンゲージメントを検出する方法600を示すフローチャートであり、これは、上記で説明された視聴者エンゲージメント評価のために使用され得る。方法600は、プロセッサ(たとえば、プロセッサ150またはリモートサーバ170中のプロセッサ)によって実装され得る。方法600は、(たとえば、メモリによって、または図1に示されているRGBセンサ130など、画像撮影デバイスから直接)ディスプレイの前の視聴エリアの画像データが与えられる、ステップ610において開始する。ステップ620において、プロセッサは、画像データから骨格フレーム(すなわち、可能性のある少なくとも1の視聴者の画像を含む画像フレーム、たとえば、図2A中の230を参照)を取得する。ステップ630において、処理ループが開始され、ここで、プロセッサは、顔認識、感情分析、およびエンゲージメント決定を含む、さらなる処理のために、各骨格フレームについて6つの個々の骨格データポイント/セットを使用する。骨格データが処理されると、方法600は、リフレッシュステップ625を介して、ステップ620における骨格フレーム取得に戻る。

20

30

【0093】

[0118]方法600におけるステップ635は判定ステップであり、ここで、プロセッサは、骨格フレーム中の選択された骨格データにおいて骨格が検出されたかどうかを判定する。検出されない場合、方法600はステップ630に戻り、ここで、新しい骨格データが処理のためにピックアップされる。少なくとも1つの骨格が検出された場合、方法600はステップ640に進み、ここで、画像データ中の視聴者の頭部エリアを識別するためのバウンディングボックスが生成される。バウンディングボックスは、たとえば、骨格情報に基づいて、たとえば、全体的骨格から頭部を識別することによって、生成され得る。

【0094】

40

[0119]再び、ステップ645は判定ステップであり、ここで、プロセッサは、バウンディングボックスが生成されたかどうか(すなわち、頭部エリアが検出されたかどうか)を判定する。画像が視聴者の全体的骨格を含むが、視聴者の頭部部分が遮られ、したがって画像にない可能性がある。この場合、方法600は、プロセッサが新しい骨格データをピックアップしたステップ630に再び戻る。バウンディングボックスが検出された場合、方法600はステップ650に進み、ここで、プロセッサは、(顔検出とも呼ばれる)第2のレベルの顔認識を行う。このステップにおいて、プロセッサは、ステップ640において生成されたバウンディングボックス内の人間の顔を検出することを試みる。顔検出は、たとえば、OpenCVにおけるHaar特徴ベースカスケード分類器(Haar Feature-based Cascade Classifier)を使用して実行され得る。さらなる情報は、その全体が

50

参照により本明細書に組み込まれる米国特許第 8, 447, 139 (B2) 号において見つけられ得る。

【0095】

[0120]ステップ655において、プロセッサは、ステップ650において顔が検出されたかどうかを判定する。検出されない場合、ステップ660において第1のレベルの顔認識が実行される。この第1のレベルの顔認識ステップは、ステップ650において実行される第2のレベルの顔認識と実質的に同様であり得る。顔検出をもう1回実行することは、顔認識技法の偶発的失敗の可能性を低減し得る。ステップ665は、ステップ655と同様の判定ステップであり、ここで、プロセッサは、顔が検出されたかどうかを判定する。

【0096】

[0121]第1のレベルの顔認識または第2のレベルの顔認識のいずれかにおいて顔が検出された場合、方法600は、顔特徴検出または顔キーポイント検出とも呼ばれる、顔ランドマーク検出を実行するためのステップ670に進む。ステップ670は、異なる顔の特徴(たとえば、眼、眉、および口の隅、鼻の先端など)のロケーションを決定するために採用される。顔ランドマーク検出のさらなる情報は、それらの全体が本明細書に組み込まれる、米国特許公開第2014/0050358(A1)号および米国特許第7,751,599(B2)号において見つけられ得る。

【0097】

[0122]ステップ672において、プロセッサは、ステップ670において顔ランドマークが検出されたかどうかを判定する。検出されない場合、方法600は、さらなる処理のために別の骨格データを選択するためのステップ630に戻る。少なくとも1つの顔ランドマークが検出された場合、判定ステップ674において、プロセッサはさらに、ステップ650における第2のレベルの顔認識において顔が検出されたかどうかを判定する。YESの場合、方法600はステップ690に進み、ここで、検出された顔が識別され(すなわち、視聴者が誰であるかを決定し)、その後、本方法はステップ680に進み、ここで、顔ランドマークに基づく顔の感情が予測される。ステップ674において、プロセッサがステップ650において顔が検出されなかったことを発見した場合、方法600は、プロセッサが視聴者の感情を推定するためのステップ680に直接進む。感情分析は、たとえば、OpenCVにおけるサポートベクターマシン(SVM)を使用して実行され得る。さらなる情報は、その全体が本明細書に組み込まれる米国特許第8,488,023号において見つけられ得る。

【0098】

[0123]一例では、図3A~図6に示されている方法は、ビデオの持続時間またはビデオの視聴者カウントにかかわらず、(TV番組および広告を含む)すべての利用可能なビデオを分析する。別の例では、図3A~図6に示されている方法は、視聴者エンゲージメントの定量分析を実行する前に、短すぎるかまたはあまりに小さい視聴者カウントを有するかのいずれかであるビデオを除外するための予備的フィルタ処理を実行する。このようにして、定量分析は、より統計的に信頼できる結果になり得る。たとえば、有数量未満の時間(たとえば、30秒未満、20秒未満、または10秒未満)の間見られたビデオが除外され得る。さらに、有限期間(たとえば、1か月、2週間、または1週間)にわたって、一定数未満の人々(たとえば、20人未満、15人未満、または10人未満)によって見られたビデオも除外され得る。

【0099】

[0124]一例では、図3A~図6に示されている方法は、ライブTV番組上で実行される。別の例では、図3A~図6に示されている方法は、録画されたTV番組上で実行される。番組のタイミングが、(たとえば、TV局のデータベースからの)その元の「フィンガークリエイションタイムスタンプ(finger creation timestamp)」から10分よりも大きくシフトしたことが認識された場合、番組は録画ウォッチングとして決定される。他の場合、番組はライブウォッチングとして決定される。

【0100】

10

20

30

40

50

[0125]コマーシャルメッセージ (CM) 効果の実験的評価

[0126]このセクションは、コマーシャルメッセージ (CM) 効果管理を考察するための、**正確な視聴データ収集および分析について説明する。「視聴可能性」と呼ばれるインデックスは、人物が「TVの前に」いるときを示す。視聴可能性インデックスは、この説明のためにおよびデータを生成する調査のために作成された。調査は、30世帯からの84人のサンプルを用いて、2週間行われた。CM曲線は、2つのシーン間の視聴可能性レートの時系列曲線を示すパターンとして定義される。シーン間のCMのパーソナル視聴レートは一定であり得るが、視聴可能性レートは変化し得る。調査結果は、CM曲線の7つのパターンがあることを示す。CMの長さおよび視聴可能性レートの変数は、CM曲線の形状に有意に寄与することがある。さらに、多項ロジットモデル (multinomial logit model) が、CM曲線を決定するのに役立つ。**

10

【0101】

[0127]この実験は、コマーシャルメッセージ (CM) と、番組と、人間の視聴態度との間の関係を調査した。実験はまた、上記で説明されたシステムおよび方法の特徴づけた。ブロードキャストタイミングおよびTV局などの番組情報と視聴態度との間の相関が統計的方法を使用して分析された。現在、日本で使用されるパーソナルオーディエンスレーティング調査は、TVのリモコン上のカラーボタンを通して人々を登録し、TV視聴の開始および終了時にその人々がカラーボタンを押したときに記録する。さらに、ピープルメータ (PM) インジケータは、TVオーディエンスが何を見たか、および誰が番組を見たかを記録する (参照により本明細書に組み込まれる、PDFフォーマットでVIDEO RESEARCH . COMウェブサイトにおいて入手可能な、Video Research Ltd.(2014):「TV rating handbook」)。ただし、このオーディエンスレーティング調査は、通常、オーディエンスレーティングが正確にキャプチャされた場合でも、集中した (focused) 視聴と何気ない (casual) 視聴とを区別することを可能にしない。

20

【0102】

[0128]HirakiおよびIto (参照により本明細書に組み込まれる、Hiraki, A.およびIto, K.(2000): Cognitive attitudes to television commercials based on eye tracking analysis combined with scenario, Japanese Journal of Human Engineering, Vol.36, 239~253ページ) は、眼球運動分析に基づく視覚情報を使用した画像認識へのCMの影響を分析するための方法を提案した。HirakiおよびItoは、再現された視聴状況の環境において、実際のCMを用いてCM視聴実験を行った。HirakiおよびItoによれば、聴覚および視覚情報は商品理解を妨げることがある。

30

【0103】

[0129]この実験では、パーソナルオーディエンスレーティングのほかに、システムによってキャプチャされる物理的存在のインジケータが、視聴態度を測定するために使用された。たとえば、CM中に、人々は座席を離れ、TVの前に座ることなしに互いに注意を向けることがある。したがって、CM中の視聴態度は、2つのインデックス、すなわちパーソナルオーディエンスレーティングと物理的存在とを使用して統計的に分析された。後者のインデックスは、本明細書では「視聴可能性 (viewability)」と呼ばれる。

【0104】

40

[0130]30世帯からの84人の視聴態度調査実験は、2014年の11月中旬から11月の終わりまで行われた。データは、14日間にわたって1日24時間採取された。

【0105】

[0131]図7は、TV702または他のディスプレイ上に示される番組または広告との、視聴エリア701中の視聴者のエンゲージメントを測定するデータ取得システム700の概略図を示す。システム700は、TV702がオンである間、視聴エリア701の画像をキャプチャする画像センサ710を含む。システム700はまた、画像センサ710からの画像データを記憶および処理し、通信ネットワークを介して未処理のおよび/または処理された画像データをサーバ (図示せず) に通信する、コンピューティングデバイス750を含む。

50

【 0 1 0 6 】

[0132]いくつかの場合には、コンピューティングデバイス750および/またはサーバは、パーソナルオーディエンスレーティングに加えて視聴可能性を測定する。視聴可能性は、「TVの前にいること」を示し、この用語は、左に70度および右に70度の間でTVの前面に顔を向けている、TVから約0.5m~約4mの距離内のオーディエンスとして定義される。一例では、視聴可能性は、1秒のレートでキャプチャされ、それは、すべてのサンプル(この場合、84)で除算された1秒間のサンプル数を示す。

【 0 1 0 7 】

[0133]図8A~図8Gは、パーソナルオーディエンスレーティングによって除算された視聴可能性の値の遷移を示す、CM曲線の7つの異なる形状を示す。この値は、TVを実際に見ている人々のパーセンテージを示すことができる。

10

【 0 1 0 8 】

[0134]CM曲線の形状の差について説明するために、データの分類およびモデル化が実行され得る。この実験において採用される分析の方法は、以下で説明される。第一に、多項ロジットモデル(たとえば、参照により本明細書に組み込まれる、Agresti, A. Categorical data analysis. John Wiley & Sons(2013)参照)が、データモデル化のために採用され得る。次いで、少なくとも、サンプルサイズ(1,065)が大きいので、非階層クラスタリングがK平均法を使用して実行され得る。次に、決定木が構成され得る。説明変数が使用され、すべてのサンプルが段階的グルーピングを使用して分類される。概して、決定木は、複数の分類ルールを木構造で表す分類モデルである。ジニ係数が不純度関数(non-purity function)として使用された。

20

【 0 1 0 9 】

[0135]これらの方法を使用してCM曲線の形状を決定するとき、分析は、CM曲線の形状を決定することに密接に関係する手法または変数をも考慮する。したがって、CMロードキャストと実質的に同時に観測される変数も含まれ得る。

【 0 1 1 0 】

[0136]1日のうちの高い視聴可能性時間範囲からのデータが使用され、これは、この実験では、18:00~24:00の6時間である。5つのTV局からのCMに対する視聴態度が分析される。TV局ごとのCM曲線の比は、図9に示されている。

【 0 1 1 1 】

[0137]分析では、CM曲線の形状は従属変数であり、図8A~図8Gに示されているように、AからGに分類される。説明変数は、CMの長さ、テレビ局、ジャンル、番組の開始からの経過時間、CMについての平均パーソナルオーディエンスレーティング、CMの平均視聴可能性レート、前のシーンについての平均パーソナルオーディエンスレーティング、前のシーンの平均視聴可能性、パーソナルオーディエンスレーティングで除算された現在のシーンの視聴可能性レート、パーソナルオーディエンスレーティングで除算された前のシーンの視聴可能性レート、ならびに日付および曜日がある。前のシーンとは、CMと前のCMとの間のシーンを指す。

30

【 0 1 1 2 】

[0138]多項ロジットモデルに基づく弁別結果は、表1に示されている。多項ロジットモデルにおける弁別率は、ランダムな弁別率よりも20%高い。CM曲線の形状がBまたはGであるとき、弁別率は特に高い。

40

【 0 1 1 3 】

[0139]このモデルでは、7つの説明変数、すなわち、CMの長さ、TV局と、番組の開始からの経過時間と、CMについての平均パーソナルオーディエンスレーティングと、視聴可能性レートと、パーソナルオーディエンスレーティングで除算されたCMの視聴可能性レートと、パーソナルオーディエンスレーティングで除算された前のシーンの視聴可能性レートとが使用される。7つの変数のうち、CMの長さおよびTV局が、弁別率に最も寄与する。

50

【表 1】

表 1. 多項ロジットモデルの結果

真 / 予測	A	B	C	D	E	F	G	合計
A	34	14	13	13	48	13	1	136
B	11	114	2	44	31	15	2	219
c	14	11	21	12	17	4	4	83
D	8	57	7	86	38	7	0	203
E	17	30	10	43	110	18	0	228
F	17	42	-	17	36	37	3	152
G	0	16	1	4	7	8	8	44
合計	101	284	54	219	287	102	18	1065
弁別率	33.66	10.14	38.89	39.27	38.33	36.27	44.44	38.50

【0114】

[0140] 7つの形状の被説明変数も、層化され得る。いくつかの異なる種類の層化が考慮され得るが、効率的な考察のために、以下の2種類の層化が比較された。

【0115】

[0141] 層化1：単調形状タイプ（C / D / E）および非単調形状タイプ（A / B / F / G）。第一に、極値を有しない単調形状タイプと、極値を有する非単調形状タイプとが、層化された。各グループに対して多項ロジットモデルが適用され、次いで、各グループについての弁別率が計算され得る。層化1の弁別結果は、表2に示されている。単調形状タイプの弁別率は59.34%であり、単調形状タイプの弁別率は51.72%であり、全体的弁別率は、53.62%である。

【0116】

[0142] 単調および非単調形状タイプを層化した後では、全体的弁別率は、層化なしの多項ロジットモデルにおける弁別率よりも15%高い。層化なしの多項ロジットモデルと比較して、CM曲線の形状間の弁別率の差は、正確に（D / E / G）および不正確に（C）決定され得る。

【0117】

[0143] 選択された説明変数は以下の通りである。単調形状タイプでは、6つの変数、すなわち、TV局と、番組の開始からの経過時間と、CMについての平均パーソナルオーディエンスレーティングと、CMの視聴可能性と、前のシーンの視聴可能性と、パーソナルオーディエンスレーティングで除算された前のシーンの視聴可能性とが選択される。非単調形状タイプでは、選択された6つの変数は、CMの長さ、TV局、番組の開始からの経過時間、CMについての平均パーソナルオーディエンスレーティング、CMの視聴可能性レート、および前のシーンの視聴可能性レートである。層化なしの多項ロジットモデルに寄与するCMの長さは、単調形状タイプでは選択されない。

10

20

30

40

50

【表 2】

表 2. 層化 1 の弁別結果

真 / 予測	A	B	C	D	E	F	G	合計
A	67	42	0	0	0	26	1	136
B	26	169	0	0	0	24	0	219
C	0	0	15	25	43	0	0	83
D	0	0	10	139	54	0	0	203
E	0	0	14	63	151	0	0	228
F	30	75	0	0	0	26	4	152
G	4	22	0	0	0	14	4	44
合計	127	308	39	227	248	90	6	1065
弁別率	52.76	54.87	38.46	61.23	60.89	28.89	66.67	53.62

【0118】

[0144]層化 2：単純な形状タイプ（A / B / C / D / E）および複雑な形状タイプ（F / G）。第二に、多くとも 1 つの極値を有する単純な形状タイプが層化され、2 つ以上の極値を有する複雑な形状タイプが層化され得る。層化 2 の弁別結果は、表 3 に示されている。単純な形状タイプの弁別率は 46.50% であり、複雑な形状タイプの弁別率は 77.55% であり、全体的弁別率は、52.21% である。

【0119】

[0145]単純な形状タイプの場合、9 つの変数、すなわち、CM の長さ、TV 局と、番組の開始からの経過時間と、CM についての平均パーソナルオーディエンスレーティングと、CM の視聴可能性レートと、前のシーンの平均パーソナルオーディエンスレーティングと、CM のパーソナルオーディエンスレーティングで除算された視聴可能性レートと、平均パーソナルオーディエンスレーティングで除算された前のシーンの視聴可能性と、日付とが選択される。さらに、複雑な形状タイプの場合、1 つの変数のみ、すなわち、TV 局が選択される。このモデルは 1 つの変数のみを有するので、すべてのサンプルが F に分類される。単純な形状タイプの場合、選択された変数は、層化なしの多項ロジットモデルのそれと同様である。

10

20

30

40

50

【表 3】

表 3. 層化 2 の弁別結果

真 / 予測	A	B	C	D	E	F	G	合計
A	39	19	13	20	45	0	0	136
B	15	121	2	46	35	0	0	219
C	12	15	23	12	21	0	0	83
D	11	so	7	103	32	0	0	203
E	22	38	10	40	118	0	0	228
F	0	0	0	0	0	152	0	152
G	0	0	0	0	0	44	0	44
合計	99	243	55	221	251	196	0	1065
弁別率	39.39	49.79	41.82	46.61	47.01	77.55	0.00	52.21

10

20

【0120】

[0146]説明変数を使用したクラスタ分析が実行され得る。クラスタ分析の弁別結果は、表 4 に示されている。弁別率は 15.77% であり、クラスタ分析とランダム選択との間の弁別率に差はない。言い換えれば、非階層クラスタ分析では、CM 曲線は分類されなかった。

【表 4】

表 4. クラスタ分析の弁別結果

真 / 予測	A	B	C	D	E	F	G	合計
A	10	21	10	14	58	14	9	136
B	22	25	19	11	116	16	10	219
C	6	10	4	11	38	10	4	83
D	17	28	6	10	110	25	7	203
E	32	29	10	13	109	28	7	228
F	11	29	7	16	76	9	4	152
G	4	7	2	3	26	1	1	44
合計	102	149	58	78	533	103	42	1065
弁別率	9.80	16.78	6.90	12.82	20.45	8.74	2.38	15.77

30

40

【0121】

[0147]図 10 は、決定木による分類モデルを示す。決定木の決定結果は、表 5 に示されている。決定木の弁別率は 40% である。表 5 から、G の弁別率は 0% であるが、D の弁

50

別率は73%程度で他のCM曲線の弁別率よりも高いことがわかる。決定木の弁別率は、層化なしの多項ロジットモデルのそれよりもわずかに高い。

【0122】

[0148]図10から、CM曲線の各形状の特性が識別され得る。形状Aは、視聴可能性レートが高いときに発生する。形状Bは、視聴可能性レートが低く、CMの長さが長いときに発生する。形状Cは、シーンの視聴可能性レートが前のシーンのそれとあまり異ならないときに発生する。形状Dは、視聴可能性レートが低く、CMの長さが短いときに発生する。形状Eは、前のシーンの視聴可能性レートが低く、CMの長さが短いときに発生する。形状Fは、シーンの視聴可能性レートは低い、前のシーンの視聴可能性レートは高いときに発生する。

【表5】

表5. 決定木の弁別結果

真 / 予測	A	B	C	D	E	F	G	合計
A	17	10	14	0	32	63	0	136
B	4	121	5	8	46	35	0	219
C	5	3	31	1	34	9	0	83
D	6	70	4	30	68	25	0	203
E	5	17	8	2	128	68	0	228
F	6	29	2	0	16	99	0	152
G	2	11	2	0	7	22	0	44
合計	45	261	66	41	331	321	0	1065
弁別率	37.78	46.36	46.97	73.17	38.67	30.84	0.00	40

【0123】

[0149]比較および考察。各方法による弁別率は、表6にまとめられる。層化1の方法は、すべての方法のうち最も高いレートを有する。ただし、被説明変数が層化されたので、コネクション全体を検証することは不可能である。

【表6】

表6. 弁別率のまとめ

多項ロジット	層化1	層化2	クラスタ分析	決定木
38.5	53.62	52.21	15.77	40

【0124】

[0150]層化なしの多項ロジットモデルの弁別率は、決定木の率とほぼ同じである。決定木は、視聴可能性レートが固定値より高いか否かによって決定されるので、直観的に理解することが困難であり、固定値は反復可能でない。したがって、CM曲線を決定するための最も好適な方法は、層化なしの多項ロジットモデルである。

【0125】

[0151]すべての方法において、CMの長さおよび視聴可能性レートの変数は、CM曲線を決定することに最も寄与する。したがって、TV視聴態度は、番組のジャンルおよびブロードキャスト時間に依存しないが、CMの長さとして現在および前のシーンの視聴可能性レートとに依存する。

【0126】

[0152]これらの5つの方法において、CMの長さおよび視聴可能性レートの変数は、CM曲線を決定することに大きく寄与する。この点について、2つのポイント、すなわち、1) CMの長さとして視聴可能性レートとの間の関係、および2) どのような状況において視聴可能性レートが高いかが考慮される。

10

【0127】

[0153]CMの長さとして視聴可能性レートとの間の関係は、図11に示されている。概して、CMの長さが短いほど、視聴可能性レートは高くなる。CMが長いほど、人々は興味をもたなくなりTVを見るのを止めることになるので、視聴可能性レートは低くなる。

【0128】

[0154]さらに、どのような状況が高い視聴可能性レートをもたらすかが調査された。(ジャンルによっては)番組が開始した後ほとんど時間が経過していないとき、視聴可能性レートは高い。表7が示すように、各ジャンルの平均視聴可能性レート間には顕著な差がある。ニュース番組の視聴可能性レートは低い、映画および音楽のそれは高い。図12は、番組の開始からの経過時間と視聴可能性レートとの間の相関を示す。図12から、番組の開始からより短い時間が経過したとき、視聴可能性レートがより高いことがわかる。

20

30

40

50

【表 7】

表 7. ジャンル別の平均視聴可能性レート

ジャンル	視聴可能性
アニメーション／特撮	0.706
スポーツ	0.668
ドキュメンタリ	0.907
ドラマ	0.807
ニュース	0.814
バラエティーショー	0.988
映画	1.252
音楽	1.359
趣味／教育	0.816
ワイドショー	0.776
全体	0.939

【 0 1 2 9 】

[0155]この実験的研究は、本発明のハードウェアおよびソフトウェア構成要素の例示的な実施形態を使用して、CMと番組と人間の視聴態度との間の関係を解明する。CM曲線を決定するための最も好適な方法は、多項ロジットモデルである。

【 0 1 3 0 】

[0156]CM中に観測され得る変数は、CM曲線とこれらの変数との間の関係を考察するために分析される。すべての採用された方法において、CMの長さおよび視聴可能性レートの変数は、CM曲線を決定することに最も寄与する。単調形状タイプの弁別率が高いので、変化なしが変化ありかにかかわらず、弁別はより容易である。言い換えれば、CM曲線の形状は、ジャンルおよび日付など、番組の特徴に関係しない。これは、CMブロードキャスト時間がより長いとき、オーディエンスは見るのが嫌になることを示す。その上、

番組の前のシーンがオーディエンスにとって面白くない場合、オーディエンスは次のCMを見ない。

【0131】

[0157]視聴者エンゲージメントデータの適用例

[0158]図13は、本明細書で説明される方法およびシステムを使用して取得されたデータの通信のシステムを示す。システム1300は、動き検知デバイスを通してTVオーディエンスパネルからキャプチャされた未処理データ1310を記憶および処理し、これは、限定はしないが、デスクトップマシンなど、コンピューティングデバイス1320に転送される。次いで、視聴者エンゲージメントを評価する方法が、たとえば、データを分析および処理するためのデスクトップマシン上で実行され得る。本方法は、分析後のデータを、(1)誰がTVを実際に見ているか(誰がオーディエンスにいるか)と、(2)オーディエンスメンバはどのくらいの頻度でTVを見るかと、(3)TV番組および広告に対するオーディエンスの反応とを決定するために使用され得る、パフォーマンスベースのTVレーティングデータに変換する。この処理されたおよび/または集約されたデータは、次いで、サーバなど、クラウド上の中央ストレージロケーション1330に転送され、ここで、限定はしないが、TV広告代理店1340、TVネットワーク1350、およびデータが有用であると発見し得る他の潜在的クライアント1360を含むサードパーティが、収集エンティティのクライアントのために特別に開発された、収集エンティティのソフトウェア、アプリケーションプログラミングインターフェース、またはウェブポータルを通して、いつでも好都合にデータにアクセスすることができる。代替的に、ハードウェア構成要素のセンサによって収集された未処理データ1310は、インターネット接続を通して直接または間接的にクラウド上の中央ストレージ1330に転送され、ここで、ソフトウェア構成要素によって分析され、関係するサードパーティ1340~1360にとって利用可能になる。サードパーティは、オプションにより、システムを通して未処理データにアクセスし得る。

【0132】

[0159]図14は、本明細書で説明されるシステムおよび方法によって取得および分析されたデータを利用することができる例示的なシステム1400の基本要素を示す。収集エンティティ1430(たとえば、TVision Insights)は、補償と引き換えにまたはボランティアにより、TV視聴率データ収集の目的で図1に示されたハードウェア構成要素の設置がパネルメンバ1410の世帯におけるテレビジョンの上に設置されることを許容する、パネルメンバ1410(たとえば、世帯のメンバ)を補償し得る。パネルメンバは、限定はしないが、クレジットカード取引データ、人口統計学的および社会的経済的情報、ソーシャルメディアアカウントログイン、ならびにタブレット、スマートフォン、および他のデバイスからのデータを含む、追加情報1420を提供するように依頼され得る。このデータは収集され、ビデオおよびIR画像は図1に示されたシステムを使用して記録され、ビデオは図2A~図6で説明された方法によって分析され得る。分析されると、ビデオを表すデータは収集エンティティ1430に送信され得、収集エンティティ1430は、次いで、広告主1440、TV局1460、TV代理店1450、および他の関係するサードパーティにデータを販売または場合によっては提供し得る。オプションにより、収集エンティティ1430は、個別の分析のために未処理の収集されたデータへのアクセスを提供し得る。開示されるビジネスモデルの一部として、収集エンティティ1430は、広告主1440に、広告主1440のTV代理店1450がこのデータを購入するように促そうという動機を与えることができる。

【0133】

[0160]図15は、視聴者エンゲージメントを評価する方法において取得されたデータに基づくビッグデータ分析および視覚化を示す。これらのモデル1500では、収集エンティティ1520(たとえば、図15に示されているTVision INSIGHTS)は、TV受像機を有する世帯1510からデータを収集することができる。引き換えに、参加する世帯1510は、収集エンティティ1520から金銭的補償(または他の利益)

10

20

30

40

50

を受け取ることができる。収集エンティティ 1520 は、次いで、特定の TV 番組または広告の有効性などの情報を導出するために、ビッグデータ分析 1530 a および視覚化技法 1530 b を使用して、参加する世帯から収集されたデータを分析する。このデータは、次いで、(顧客 1540 と総称される) 広告主、広告代理店、TV 局、あるいは他のコンテンツプロバイダまたはプロモーターに、番組の有効性を向上させるように指示するために、提供され得る。一例では、顧客 1540 は、月ごとに月額料金で、収集エンティティ 1520 にこのデータサービスを申し込むことができる。別の例では、顧客 1540 は、収集エンティティ 1520 から、特定のビデオ(たとえば、キャンペーンビデオ、スポーツイベント中の特別な広告など)に関係するデータを買うことができる。

【0134】

[0161] 図 16 は、視聴者エンゲージメントデータ収集に参加する個人および世帯(TV オーディエンス)からの追加情報 1600 の収集の例を示す。TV オーディエンスは、関係するサードパーティに有用な国および/または地域の層を表し得る。収集エンティティは、ビデオデータ 1610 および人口統計学的情報を収集し、システムによって集められ TV 視聴率に関する方法によって分析されたデータとともにパッケージ化して、この情報を有償で顧客に提供することができる。TV オーディエンスから収集され得る情報の例は、限定はしないが、特に、TWITTER(登録商標)、Instagram、FACEBOOK(登録商標)など、ソーシャルメディアプロフィール 1620 を通して取得され得るすべての情報を含む。情報は、(テレビジョンオーディオと世帯中の個人から発せられた会話などのオーディオの両方を含む)システムから取得されたビデオデータおよびオーディオデータ 1640 と、スマートフォンおよびタブレット検索傾向、インターネット検索履歴、電子メールアカウント情報を含むマルチスクリーンデータ 1630 と、クレジットカード取引データ 1650 とをさらに含むことができる。このリストは網羅的なものではなく、限定するものとして解釈されるべきではない。

【0135】

[0162] 広告主にとって有益である、個々の層の先例のない測定値を含む、収集された情報およびデータは、収集エンティティが、TV 広告の影響を正確に評価することを可能にする。広告主は、どの広告枠が広告主のターゲットオーディエンスに対する最良適合であるかを決定するために、データを使用することができる。また、メッセージは、オーディエンスのタイプに一層関連するものになり得、購買行動に効果的につながり、広告主にとっての投資収益率(ROI)を高めることができる。

【0136】

[0163] TV ネットワークも、その TV 番組のより正確なレーティング、オーディエンスタイプ、反応、および予測広告枠価値を収集することが可能になるので、開示される発明から恩恵を受けることができる。これは、どの広告枠が特定のターゲット層に対して最も高い価値を有することになるかを決定することに加えて、TV ネットワークがその番組をオーディエンスのタイプにより良く適合するように改善し、人気がない番組をなくすことを可能にする。データはまた、番組および広告の比較評価のために、同じまたは異なる時間枠において、複数のチャンネルにわたって番組を比較するために使用され得る。同様に、TV オーディエンスデータおよび行動は、ストリーミングコンテンツに対する所与の番組時間枠について、収集および比較され得る。TV パイロット番組も、エピソードを発注する前に、システムを使用して評価され得る。

【0137】

[0164] 結論

[0165] 様々な本発明の実施形態が本明細書で説明および図示されているが、機能を実行するための、ならびに/あるいは結果および/または本明細書で説明される利点のうちの 1 つまたは複数を得るための、様々な他の手段および/または構造を、当業者は容易に想定され、そのような変形および/または変更の各々は、本明細書で説明される本発明の実施形態の範囲内であると見なされる。より一般的には、本明細書で説明されるすべてのパラメータ、寸法、材料、および構成が例示的なものであることを意味し、実際のパラメー

10

20

30

40

50

タ、寸法、材料、および/または構成が、本発明の教示が使用される特定の1つまたは複数の適用例に依存することを、当業者は容易に諒解されよう。当業者は、通常の実験だけを使用して、本明細書で説明される特定の発明の実施形態の多くの均等物を認識し、または確認することが可能であろう。したがって、上記の実施形態は単に例として提示され、添付の特許請求の範囲およびその均等物の範囲内で、本発明の実施形態は、特に説明および請求された以外の他の方法で実施され得ることを理解されたい。本開示の発明の実施形態は、本明細書で説明されるそれぞれ個々の特徴、システム、物品、材料、キット、および/または方法を対象とするものである。さらに、2つまたはそれ以上のそのような特徴、システム、物品、材料、キット、および/または方法の任意の組合せは、そのような特徴、システム、物品、材料、キット、および/または方法が相互に矛盾しない場合、本開示の発明の範囲内に含まれる。

10

【0138】

[0166]上記で説明された実施形態は、多数の方法のいずれかで実装され得る。たとえば、本明細書で開示される技術を設計および製作する実施形態は、ハードウェア、ソフトウェア、またはそれらの組合せを使用して実装され得る。ソフトウェアで実装される時、単一のコンピュータ中で提供されるか複数のコンピュータの間で分散されるかにかかわらず、ソフトウェアコードは任意の好適なプロセッサまたはプロセッサの集合上で実行され得る。

【0139】

[0167]さらに、コンピュータは、ラックマウント式コンピュータ、デスクトップコンピュータ、ラップトップコンピュータ、またはタブレットコンピュータなど、いくつかの形式のいずれかで実施され得ることを諒解されたい。さらに、コンピュータは、携帯情報端末(PDA)、スマートフォン、あるいは他の好適なポータブルまたは固定電子デバイスを含む、概してコンピュータとして見なされないが好適な処理能力をもつデバイス中に埋め込まれ得る。

20

【0140】

[0168]また、コンピュータは1つまたは複数の入出力デバイスを有し得る。これらのデバイスは、特に、ユーザインターフェースを提示するために使用され得る。ユーザインターフェースを提供するために使用され得る出力デバイスの例は、出力の視覚提示のためのプリンタまたはディスプレイスクリーン、および出力の可聴提示のためのスピーカまたは他の音発生デバイスを含む。ユーザインターフェースのために使用され得る入力デバイスの例は、キーボード、ならびにマウス、タッチパッド、およびデジタル化タブレットなど、ポインティングデバイスを含む。別の例として、コンピュータは、音声認識を通して、または他の可聴フォーマットで入力情報を受信し得る。

30

【0141】

[0169]そのようなコンピュータは、エンタープライズネットワークなど、ローカルエリアネットワークまたはワイドエリアネットワーク、およびインテリジェントネットワーク(IN)またはインターネットを含む、任意の好適な形式で1つまたは複数のネットワークによって相互接続され得る。そのようなネットワークは、任意の好適な技術に基づき得、任意の好適なプロトコルに従って動作し得、ワイヤレスネットワーク、ワイヤードネットワークまたは光ファイバーネットワークを含み得る。

40

【0142】

[0170]本明細書で概説される様々な方法またはプロセスは、様々なオペレーティングシステムまたはプラットフォームのいずれか1つを採用する、1つまたは複数のプロセッサ上で実行可能なソフトウェアとしてコーディングされ得る。さらに、そのようなソフトウェアは、いくつかの好適なプログラミング言語および/あるいはプログラミングまたはスクリプティングツールのいずれかを使用して記述され得、また、フレームワークまたは仮想マシン上で実行される、実行可能機械語コードまたは中間コードとしてコンパイルされ得る。

【0143】

50

[0171]この点において、様々な本発明の概念は、1つまたは複数のコンピュータまたは他のプロセッサ上で実行されたとき、上記で説明された本発明の様々な実施形態を実装する方法を実行する1つまたは複数のプログラムで符号化された、コンピュータ可読記憶媒体（または複数のコンピュータ可読記憶媒体）（たとえば、コンピュータメモリ、1つまたは複数のフロッピー（登録商標）ディスク、コンパクトディスク、光ディスク、磁気テープ、フラッシュメモリ、フィールドプログラマブルゲートアレイまたは他の半導体デバイス中の回路構成、あるいは他の非一時的媒体または有形コンピュータ記憶媒体）として具現化され得る。コンピュータ可読媒体またはメディアは、記憶された1つまたは複数のプログラムが、上記で説明された本発明の様々な態様を実装するために、1つまたは複数の異なるコンピュータまたは他のプロセッサ上にロードされ得るように、トランスポート可能であり得る。

10

【0144】

[0172]「プログラム」または「ソフトウェア」という用語は、本明細書では、上記で説明された実施形態の様々な態様を実装するようにコンピュータまたは他のプロセッサをプログラムするために採用され得る、任意のタイプのコンピュータコードまたはコンピュータ実行可能命令のセットを指すために、一般的な意味で使用される。さらに、一態様によれば、実行されたとき本発明の方法を実行する1つまたは複数のコンピュータプログラムは、単一のコンピュータまたはプロセッサ上に存在する必要はないが、本発明の様々な態様を実装するために、いくつかの異なるコンピュータまたはプロセッサの間においてモジュール様式で分散され得ることを諒解されたい。

20

【0145】

[0173]コンピュータ実行可能命令は、プログラムモジュールなど、1つまたは複数のコンピュータまたは他のデバイスによって実行される多くの形式であり得る。概して、プログラムモジュールは、特定のタスクを実行するか、または特定の抽象データ型を実装する、ルーチン、プログラム、オブジェクト、構成要素、データ構造などを含む。一般に、プログラムモジュールの機能は、様々な実施形態において必要に応じて組み合わせられるかまたは分散され得る。

【0146】

[0174]また、データ構造は、任意の好適な形式でコンピュータ可読媒体に記憶され得る。説明を簡単にするために、データ構造は、データ構造における場所を通して関係付けられるフィールドを有するように示されることがある。そのような関係は、フィールド間の関係を伝達するコンピュータ可読媒体における場所をフィールドの格納に割り当てることによって、同様に達成され得る。ただし、データ要素間の関係を確立するポインタ、タグ、または他の機構の使用によるものを含む、任意の好適な機構が、データ構造のフィールド中の情報間の関係を確立するために使用され得る。

30

【0147】

[0175]また、様々な本発明の概念は、その例が与えられている1つまたは複数の方法として実施され得る。本方法の一部として実行される行為は、任意の好適な方法で順序付けられ得る。したがって、例示的な実施形態において連続的な行為として示されている場合であっても、いくつかの行為を同時に実行することを含み得る、図示されたものとは異なる順序で行為が実行される実施形態が構成され得る。

40

【0148】

[0176]本明細書で定義および使用されるすべての定義は、辞書の定義、参照により組み込まれる文書中の定義、および/または定義された用語の通常の意味を統制することを理解されたい。

【0149】

[0177]本明細書および特許請求の範囲において、本明細書で使用される不定冠詞「a」および「an」は、そうでないことが明確に示されていない限り、「少なくとも1つ」を意味することを理解されたい。

【0150】

50

[0178]本明細書および特許請求の範囲において、本明細書で使用される「および/または」という句は、そのように結合された要素、すなわち、いくつかの場合には結合して存在し、他の場合には分離して存在する要素の「いずれかまたは両方」を意味することを理解されたい。「および/または」を用いて列挙される複数の要素は、同じように、すなわち、そのように結合された要素のうちの「1つまたは複数」と解釈されたい。「および/または」節によって具体的に特定された要素以外の他の要素が、具体的に特定されたそれらの要素に関係するか関係しないかにかかわらず、随意に存在し得る。したがって、非限定的な例として、「Aおよび/またはB」への言及は、「備える (comprising)」などの非限定的用語とともに使用されるとき、一実施形態では (B以外の要素を随意に含む) Aのみを指し、別の実施形態では (A以外の要素を随意に含む) Bのみを指し、また別の実施形態では (他の要素を随意に含む) AとBの両方を指すことができる、などである。

10

【0151】

[0179]本明細書および特許請求の範囲において、本明細書で使用される「または」は、上記で定義された「および/または」と同じ意味を有することを理解されたい。たとえば、リスト中の項目を分離するとき、「または」または「および/または」は、包含的なものであり、すなわち、いくつかの要素または要素のリスト、および随意に、リストに載っていない追加の項目のうちの、2つ以上も含む、少なくとも1つを含むこととして解釈されるものとする。「のうちの1つのみ」または「のうちの厳密に1つ」、あるいは特許請求の範囲で使用されるとき、「からなる (consisting of)」など、そうでないことが明確に示されている用語のみが、いくつかの要素または要素のリストのうちの厳密に1つの要素を含むことを指すことになる。一般に、本明細書で使用される「または」という用語は、「いずれか」、「のうちの1つ」、「のうちの1つのみ」、または「のうちの厳密に1つ」など、排他的用語に先行されるとき、排他的代替 (すなわち、「一方または他方であるが両方ではない」) を示すものとして解釈されるものとする。特許請求の範囲で使用されるとき、「から本質的になる (consisting essentially of)」は、特許法の分野で使用されるその通常の意味を有するものとする。

20

【0152】

[0180]本明細書および特許請求の範囲において、本明細書で使用される「少なくとも1つ」という句は、1つまたは複数の要素のリストに関して、要素のリスト中の要素のうちの1つまたは複数から選択される少なくとも1つの要素を意味するが、必ずしも要素のリスト内に具体的にリストされたあらゆる要素のうちの少なくとも1つを含むとは限らず、要素のリスト中の要素のいかなる組合せも除外しないことを理解されたい。この定義はまた、「少なくとも1つ」という句が指す要素のリスト内で具体的に特定された要素以外の要素が、具体的に特定されたそれらの要素に関係するか関係しないかにかかわらず、随意に存在することを可能にする。したがって、非限定的な例として、「AおよびBのうちの少なくとも1つ」(または等価的に「AまたはBのうちの少なくとも1つ」、または等価的に「Aおよび/またはBのうちの少なくとも1つ」) は、一実施形態では、Bが存在せず (B以外の要素を随意に含む)、2つ以上を随意に含む、少なくとも1つのAを指し、別の実施形態では、Aが存在せず (A以外の要素を随意に含む)、2つ以上を随意に含む、少なくとも1つのBを指し、また別の実施形態では、(他の要素を随意に含む) 2つ以上を随意に含む、少なくとも1つのA、および2つ以上を随意に含む、少なくとも1つのBを指すことができる、などである。

30

40

【0153】

[0181]特許請求の範囲において、ならびに上記の明細書において、「備える」、「含む (including)」、「搬送する (carrying)」、「有する (having)」、「含んでいる (containing)」、「伴う (involving)」、「保持する (holding)」、「から構成される (composed of)」など、すべての移行句は、非限定的なもの、すなわち、限定はしないが含むことを意味するものであることを理解されたい。「からなる」および「から本質的になる」という移行句のみが、それぞれ、米国特許庁特許審査基準、セクション 2111.03に記載されている、限定的または半限定的な移行句であるものとする。

50

以下に、出願当初の特許請求の範囲に記載された発明を付記する。

[C 1]

ディスプレイ上で再生しているビデオとの視聴者エンゲージメントを定量化するためのシステムであって、

前記ディスプレイの前の視聴エリアを撮像するように配設された、前記視聴エリアの画像データを取得するための少なくとも1つのカメラと、

前記ディスプレイに近接して配設された、前記ディスプレイに結合されたスピーカによって発せられたオーディオデータを取得するためのマイクロフォンと、

前記少なくとも1つのカメラおよび前記マイクロフォンに動作可能に結合された、プロセッサ実行可能命令を記憶するためのメモリと、

前記少なくとも1つのカメラ、前記マイクロフォン、および前記メモリに動作可能に結合されたプロセッサとを備え、ここにおいて、前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記少なくとも1つのカメラからの前記画像データ、および前記マイクロフォンからの前記オーディオデータを受信することと、

前記オーディオデータに少なくとも一部基づいて、前記ディスプレイ上に表示された前記ビデオの識別情報を決定することと、

前記画像データに少なくとも一部基づいて、前記視聴エリア中に存在する人々の第1の数、および前記視聴エリア中の前記ビデオにエンゲージしている人々の第2の数を推定することと、

人々の前記第1の数および人々の前記第2の数に少なくとも一部基づいて、前記ビデオの前記視聴者エンゲージメントを定量化することと

を行う、システム。

[C 2]

前記ビデオは、セットトップボックスを介して提供されるテレビ番組を含み、前記プロセッサは、前記セットトップボックスに接続されない、上記C 1に記載のシステム。

[C 3]

前記少なくとも1つのカメラは、可視カメラと赤外線カメラとを含み、前記画像データは、前記可視カメラによって取得された第1の画像と、前記赤外線カメラによって取得された第2の画像とを含む、上記C 1に記載のシステム。

[C 4]

前記プロセッサ実行可能命令の実行時に、前記プロセッサはさらに、

前記第1の画像から人々の第1の未処理の数、および前記第2の画像から人々の第2の未処理の数を推定することと、

前記第1の未処理の数または前記第2の未処理の数のうちの少なくとも一方における、可能性のあるエラーを検出するために、前記第1の未処理の数を前記第2の未処理の数と比較することと

を行う、上記C 3に記載のシステム。

[C 5]

前記少なくとも1つのカメラは、実質的に毎秒1フレーム以上のフレームレートで前記画像データを取得する、上記C 1に記載のシステム。

[C 6]

前記プロセッサは、身体骨格検出に基づいて、前記視聴エリア中に存在する人々の前記第1の数を推定する、上記C 1に記載のシステム。

[C 7]

前記プロセッサは、視線追跡に基づいて、前記ビデオにエンゲージしている人々の前記第2の数を推定する、上記C 1に記載のシステム。

[C 8]

前記マイクロフォンは、約0.1 Hzの取得レートで前記オーディオデータを取得する、上記C 1に記載のシステム。

10

20

30

40

50

[C 9]

前記プロセッサは、オーディオ信号フィンガープリンティングを使用して、前記ビデオの前記識別情報を決定する、上記 C 1 に記載のシステム。

[C 1 0]

前記プロセッサは、

前記ビデオについての注意レートを推定することによって前記視聴者エンゲージメントを定量化し、前記注意レートは、所与の時間期間にわたる、前記視聴エリア中の人々の前記第 1 の数に対する、前記ビデオにエンゲージしている人々の前記第 2 の数の比を表す、上記 C 1 に記載のシステム。

[C 1 1]

前記ビデオは、複数のビデオ中の一意のビデオであり、前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記画像データと前記視聴エリア中に潜在的に存在する人々に関する人口統計学的情報とに基づいて、視聴者カウントとポジティブ持続時間比とを推定し、前記視聴者カウントは、各一意のビデオにエンゲージしている人々の前記第 2 の数を表し、前記ポジティブ持続時間比は、前記一意のビデオの持続時間に対する、前記視聴エリア中の人々が前記一意のビデオを見ることによって費やされた合計時間の比を表す、

上記 C 1 0 に記載のシステム。

[C 1 2]

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記画像データに少なくとも一部基づいて、前記視聴エリア中に存在する各人物の識別情報を決定することと、

識別された各人物についての前記視聴者エンゲージメントを定量化することとを行う、上記 C 1 に記載のシステム。

[C 1 3]

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

人々の前記第 1 の数および人々の前記第 2 の数を前記メモリに記憶することと、

前記画像データを消去および/または上書きすることと

を行う、上記 C 1 に記載のシステム。

[C 1 4]

前記プロセッサに動作可能に結合された、人々の前記第 1 の数と人々の前記第 2 の数とをリモートサーバに送信するためのネットワークインターフェース

をさらに備える、上記 C 1 に記載のシステム。

[C 1 5]

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記視聴エリア中に存在する各人物の感情を推定する

上記 C 1 に記載のシステム。

[C 1 6]

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記オーディオデータに少なくとも一部基づいて、前記複数のビデオ中のあらかじめ決定されたビデオが前記ディスプレイ上に表示されているかどうかを決定し、ここにおいて、前記視聴者エンゲージメントを定量化することは、前記あらかじめ決定されたビデオが表示されているかどうかに少なくとも一部基づく、

上記 C 1 に記載のシステム。

[C 1 7]

前記プロセッサ実行可能命令の実行時に、前記プロセッサは、

前記画像データから前記視聴エリア中の各人物についての人口統計学的情報を推定すること

を行う、上記 C 1 に記載のシステム。

[C 1 8]

10

20

30

40

50

前記人口統計学的情報は、年齢、性別、民族グループ、および顔の表情を含む、上記 C 17 に記載のシステム。

[C 19]

ディスプレイ上に示されるビデオとの視聴者エンゲージメントを定量化する方法であって、

少なくとも1つのカメラを用いて、前記ビデオが前記ディスプレイ上に示されている間、前記ディスプレイの前の視聴エリアの画像を取得することと、

マイクロフォンを用いて、前記ディスプレイに結合されたスピーカによって発せられた前記ビデオのサウンドトラックを表すオーディオデータを取得することと、

前記少なくとも1つのカメラおよびプロセッサに動作可能に結合された前記プロセッサを用いて、前記オーディオデータに少なくとも一部基づいて、前記ビデオの識別情報を決定することと、

前記プロセッサを用いて、前記画像データに少なくとも一部基づいて、前記ビデオが前記ディスプレイ上に示されている間の前記視聴エリア中に存在する人々の第1の数、および前記視聴エリア中の前記ビデオにエンゲージしている人々の第2の数を推定することと、

前記プロセッサによって、前記ビデオの前記識別情報、人々の前記第1の数、および人々の前記第2の数をリモートサーバに送信することと

を備える、方法。

[C 20]

前記画像を取得することは、可視カメラを使用して前記視聴エリアの第1の画像を取得することと、赤外線 (I R) カメラを使用して前記視聴エリアの第2の画像を取得することとを含む、上記 C 19 に記載の方法。

[C 21]

前記視聴エリア中の人々の前記第1の数を推定することは、

前記第1の画像データから人々の第1の未処理の数、および前記第2の画像データから人々の第2の未処理の数を推定することと、

前記第1の未処理の数または前記第2の未処理の数のうちの少なくとも一方における、可能性のあるエラーを検出するために、前記第1の未処理の数を前記第2の未処理の数と比較することと

を含む、上記 C 20 に記載の方法。

[C 22]

前記画像データを取得することは、実質的に毎秒 20 フレーム以上のフレームレートで前記視聴エリアの画像を取得することを含む、上記 C 19 に記載の方法。

[C 23]

前記オーディオデータを取得することは、約 0 . 1 H z の取得レートで前記オーディオデータを取得することを含む、上記 C 19 に記載の方法。

[C 24]

前記ビデオの前記識別情報を決定することは、オーディオ信号フィンガープリンティングに基づく、上記 C 19 に記載の方法。

[C 25]

前記視聴エリア中に存在する人々の前記第1の数を推定することは、身体骨格検出に基づく、上記 C 19 に記載の方法。

[C 26]

前記少なくともオンビデオにエンゲージしている人々の前記第2の数を推定することは、視線追跡に基づく、上記 C 19 に記載の方法。

[C 27]

複数の世帯中の各家世帯において、人々の前記第1の数および人々の前記第2の数に少なくとも一部基づいて、前記ビデオの前記視聴者エンゲージメントを定量化することをさらに備える、上記 C 19 に記載の方法。

[C 28]

10

20

30

40

50

前記視聴者エンゲージメントを定量化することは、

前記ビデオについての注意レートを推定することであって、前記注意レートは、前記視聴エリア中の人々の前記第 1 の数に対する、前記ビデオにエンゲージしている人々の前記第 2 の数の比を表す、推定することと、

複数のビデオ中の各一意のビデオについて、前記複数のビデオ中の前記ビデオの前記注意レートに基づいて注意インデックスを決定することと

を含む、上記 C 2 7 に記載の方法。

[C 2 9]

前記ビデオは、複数のビデオ中の一意のビデオであり、前記方法は、

前記画像データと前記複数の世帯中の各世帯に関する人口統計学的情報とに基づいて、
10
視聴者カウントとポジティブ持続時間比とを推定することをさらに備え、前記視聴者カウントは、各一意のビデオにエンゲージしている人々の前記第 2 の数を表し、前記ポジティブ持続時間比は、前記一意のビデオの持続時間に対する、前記複数の世帯中の人々が前記一意のビデオを見ることによって費やされた合計時間の比を表す、

上記 C 2 8 に記載の方法。

[C 3 0]

前記画像データに少なくとも一部基づいて、前記視聴エリア中に存在する各人物の識別
情報を決定することをさらに備え、

20
ここにおいて、前記ビデオの前記視聴者エンゲージメントを定量化することは、識別された各人物についての前記視聴者エンゲージメントを定量化することを含む、

上記 C 2 7 に記載の方法。

[C 3 1]

人々の前記第 1 の数と人々の前記第 2 の数とをリモートサーバに送信することをさらに
備え、ここにおいて、前記視聴者エンゲージメントを定量化することは、前記リモートサ
ーバにおいて行われる、上記 C 2 7 に記載の方法。

[C 3 2]

前記オーディオデータに少なくとも一部基づいて、前記複数のビデオ中のあらかじめ決
定されたビデオが前記ディスプレイ上に表示されているかどうかを決定することをさらに
備え、ここにおいて、前記視聴者エンゲージメントを定量化することは、前記あらかじめ
決定されたビデオが表示されているかどうか少なくとも一部基づく、

30
上記 C 2 7 に記載の方法。

[C 3 3]

人々の前記第 1 の数および人々の前記第 2 の数を、前記プロセッサに動作可能に結合さ
れたメモリに記憶することと、

前記画像データを消去および/または上書きすることと

をさらに備える、上記 C 1 9 に記載の方法。

[C 3 4]

前記視聴エリア中に存在する各人物の感情を推定することをさらに備える、上記 C 1 9
に記載の方法。

[C 3 5]

40
前記画像データから前記視聴エリア中の各人物についての人口統計学的情報を推定する
こと

をさらに備える、上記 C 1 9 に記載の方法。

[C 3 6]

前記人口統計学的情報を推定することは、年齢、性別、民族グループ、および顔の表情
を推定することを含む、上記 C 3 5 に記載の方法。

[C 3 7]

ディスプレイ上で再生しているビデオとの視聴者エンゲージメントを評価するためのシ
ステムであって、前記ディスプレイは、前記ビデオのサウンドトラックを発するスピーカ
に結合され、前記システムは、

第1のサンプルレートで、前記ビデオが前記ディスプレイ上で再生している間、前記ディスプレイの前の視聴エリアの可視画像を取得するための可視カメラと、

前記第1のサンプルレートで、前記ビデオが前記ディスプレイ上で再生している間、前記ディスプレイの前の前記視聴エリアの赤外線画像を取得するための赤外線カメラと、

前記ディスプレイに近接して配設された、前記第1のサンプルレートよりも低い第2のサンプルレートで、前記ビデオが前記ディスプレイ上で再生している間、前記スピーカによって発せられた前記サウンドトラックのサンプルを取得するためのマイクロフォンと、

前記可視カメラ、前記赤外線カメラ、および前記マイクロフォンに動作可能に結合されたプロセッサであって、

(i) 前記サウンドトラックの前記サンプルに基づいて前記ビデオを識別すること、

(i i) 前記可視画像および前記赤外線画像に基づいて、前記ビデオが前記ディスプレイ上で再生している間の前記視聴エリア中の人々の数、および前記ビデオにエンゲージしている人々の数を推定すること、ならびに

(i i i) 前記サウンドトラックの前記サンプル、前記可視画像、および前記赤外線画像を上書きおよび/または消去すること

を行うためのプロセッサと、

前記プロセッサに動作可能に結合された、前記ビデオの識別情報、前記ビデオが前記ディスプレイ上で再生している間の前記視聴エリア中の人々の前記数、および前記ビデオにエンゲージしている人々の前記数の表現を記憶するためのメモリと、

前記プロセッサに動作可能に結合された、前記表現をサーバに送信するためのネットワークインターフェースと

を備える、システム。

[C 3 8]

複数のビデオ中の一意のビデオについての視聴者エンゲージメントを定量化する方法であって、

複数の世帯中の各世帯において、ディスプレイの前の視聴エリアの画像データを取得することと、

前記ディスプレイが前記複数のビデオ中のビデオを示しているかどうかを決定することと、

前記複数のビデオ中の各一意のビデオについて、前記画像データおよび前記複数の世帯中の各世帯に関する人口統計学的情報に基づいて、(i) 視聴レートおよび(i i) ウォッチングレートを推定することであって、前記視聴レートは、ビデオを示しているディスプレイの合計数に対する、前記視聴エリア中の人々の合計数の比を表し、前記ウォッチングレートは、前記複数の世帯中の人々の合計数に対する、ディスプレイがビデオを示している世帯中の人々の合計数の比を表す、推定することと、

前記複数のビデオ中の各一意のビデオについて、前記視聴レートおよび前記ウォッチングレートに基づいて、視聴可能性インデックスを決定することと

を備える、方法。

[C 3 9]

前記複数のビデオ中の各一意のビデオについて、前記画像データおよび前記複数の世帯中の各世帯に関する人口統計学的情報に基づいて、(i i i) 視聴者カウントおよび(i v) ポジティブ持続時間比を推定することであって、前記視聴者カウントは、各一意のビデオにエンゲージしている人々の合計数を表し、前記ポジティブ持続時間比は、前記一意のビデオの持続時間に対する、前記複数の世帯中の人々が前記一意のビデオを見ることによって費やされた合計時間の比を表す、推定することと、

前記視聴者カウントおよび前記ポジティブ持続時間比に基づいて、前記視聴可能性インデックスを重み付けすることと

をさらに備える、上記 C 3 8 に記載の方法。

[C 4 0]

前記複数のビデオ中の前記一意のビデオにわたる前記視聴可能性インデックスを正規化

10

20

30

40

50

することをさらに備える、上記C 3 9に記載の方法。

[C 4 1]

前記画像データを取得することは、光学カメラを使用して前記視聴エリアの第 1 の画像を取得することと、赤外線 (I R) カメラを使用して前記視聴エリアの第 2 の画像を取得することを含む、上記C 3 8に記載の方法。

[C 4 2]

前記ディスプレイが前記ビデオを示しているかどうかを決定することは、信号フィンガープリンティング技法を介した前記視聴エリアのオーディオデータに少なくとも一部基づく、上記C 3 8に記載の方法。

[C 4 3]

前記視聴レートと前記ウォッチングレートとをリモートサーバに送信することをさらに備え、ここにおいて、前記視聴可能性インデックスは、前記リモートサーバによって推定される、

上記C 3 8に記載の方法。

【 図 面 】

【 図 1 】

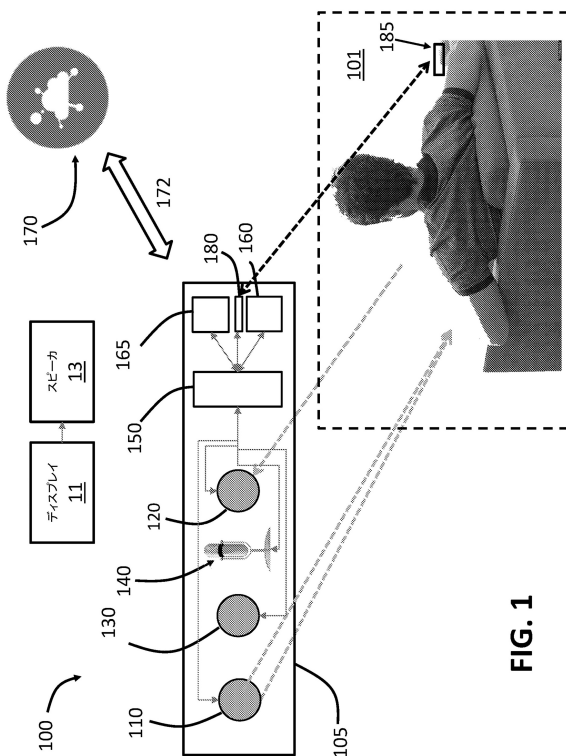


FIG. 1

【 図 2 A 】

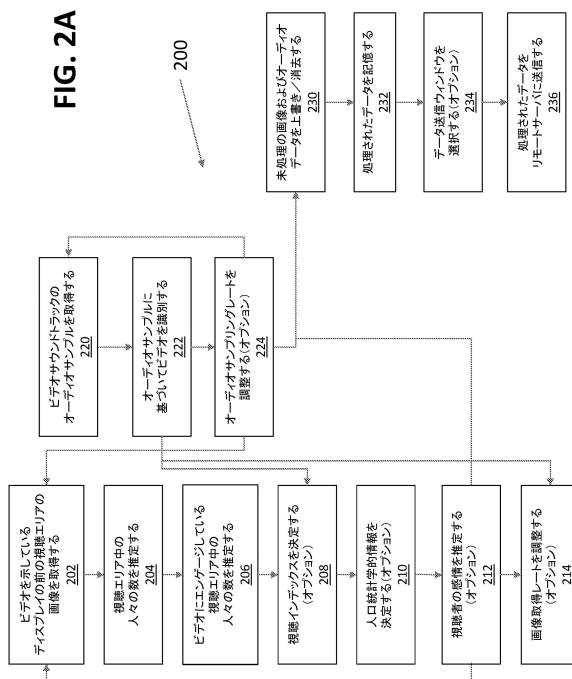


FIG. 2A

10

20

30

40

50

【図 2 B】

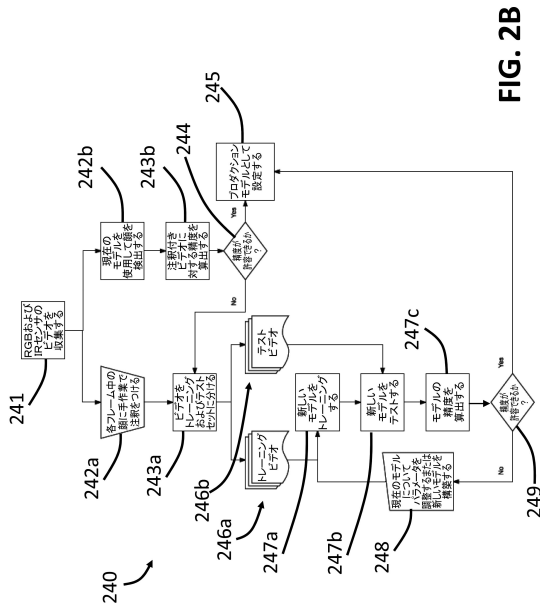


FIG. 2B

【図 3 A】

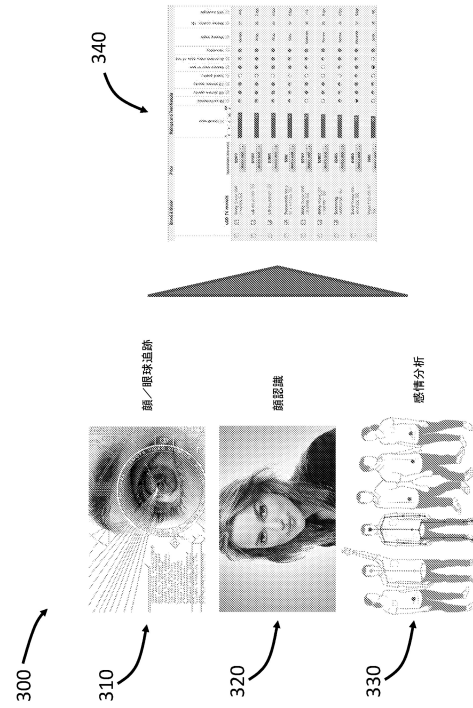


FIG. 3A

10

20

【図 3 B】

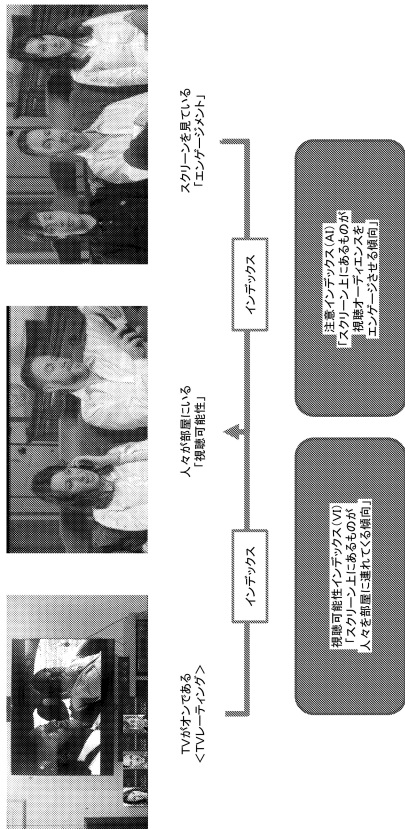


FIG. 3B

【図 4 A】

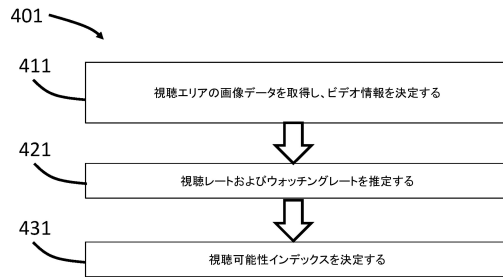


FIG. 4A

30

40

50

【 図 4 B 】

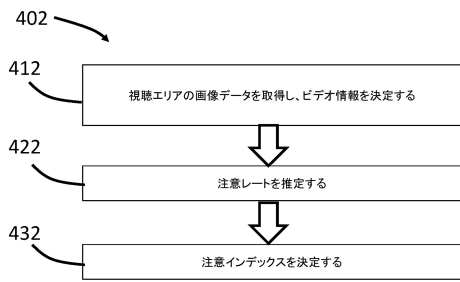


FIG. 4B

【 図 5 】

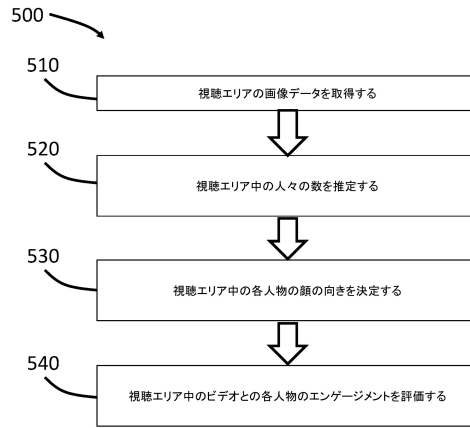


FIG. 5

【 図 6 】

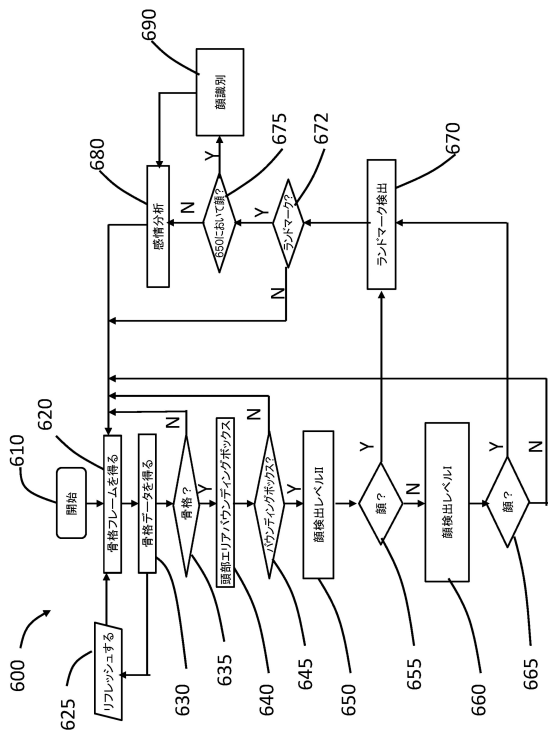


FIG. 6

【 図 7 】

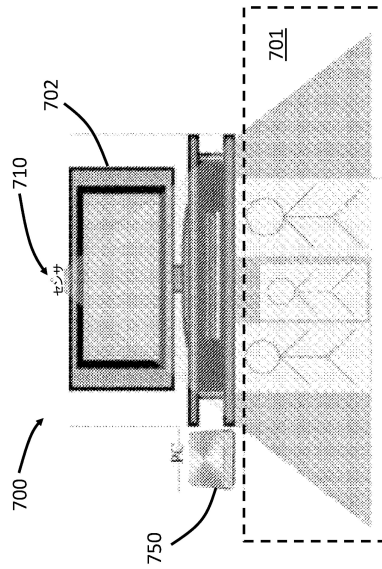


FIG. 7

10

20

30

40

50

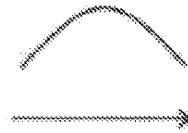
【 図 8 A 】



A. 下向き凸形

FIG. 8A

【 図 8 B 】



B. 上向き凸形

FIG. 8B

【 図 8 C 】



C. 平行

FIG. 8C

【 図 8 D 】



D. 単調に増加する

FIG. 8D

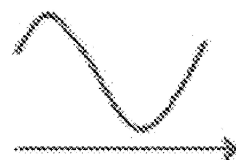
【 図 8 E 】



E. 単調に減少する

FIG. 8E

【 図 8 F 】



F. 規則性なし

FIG. 8F

10

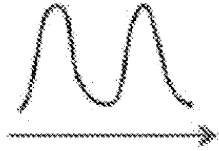
20

30

40

50

【 8 G 】



G. 2つの山

FIG. 8G

【 9 】

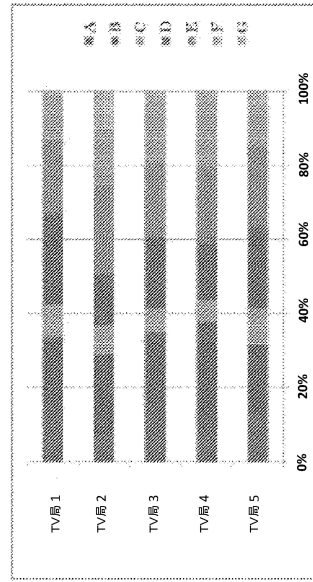


FIG. 9

10

20

【 10 】

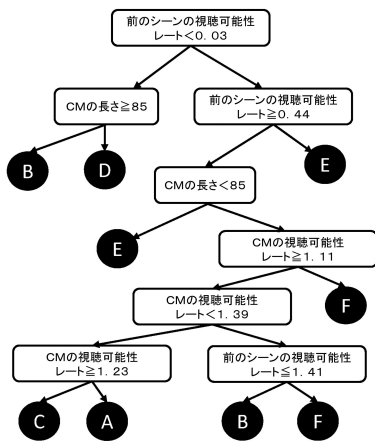


FIG. 10

【 11 】

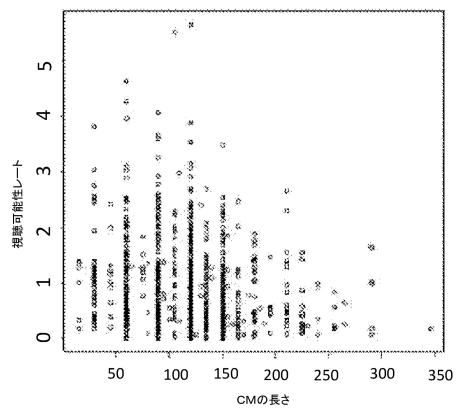


FIG. 11

30

40

50

【図 1 2】

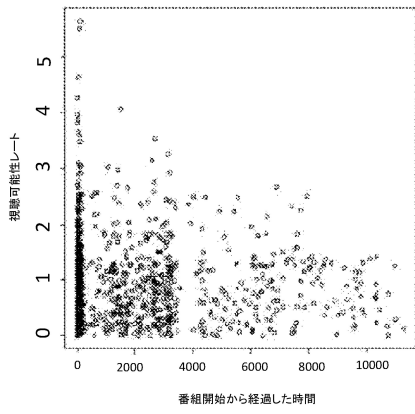


FIG. 12

【図 1 3】

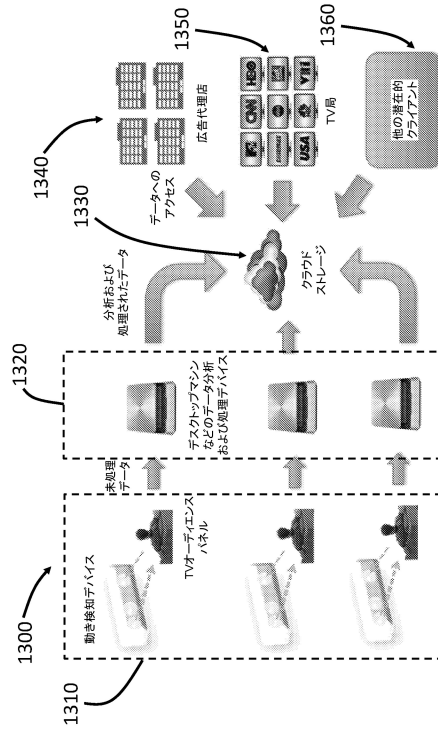


FIG. 13

10

20

【図 1 4】

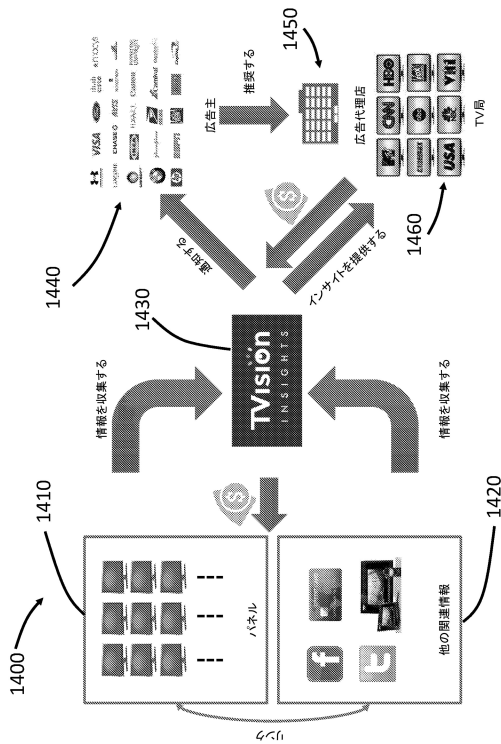


FIG. 14

【図 1 5】

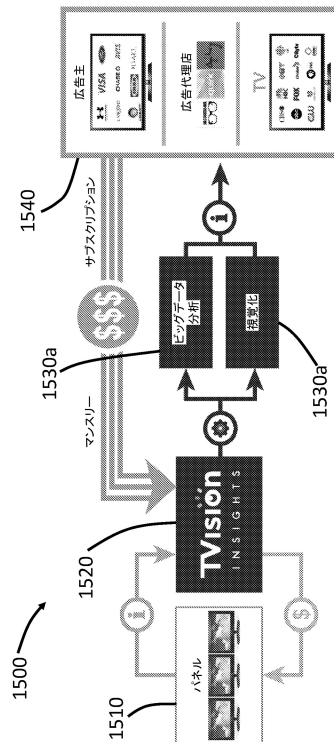


FIG. 15

30

40

50

【 図 16 】

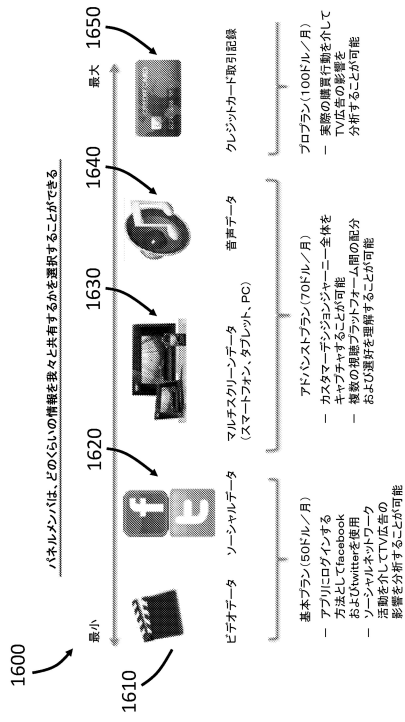


FIG. 16

10

20

30

40

50

フロントページの続き

- ツウェンティーナインス・ストリート 34 - 19、ナンバー 5ディー
(72)発明者 フ、ユン
アメリカ合衆国、マサチューセッツ州 02446、ブロックライン、ビーコン・ストリート 14
54、アパートメント 641
- 審査官 川中 龍太
- (56)参考文献 米国特許出願公開第2014/0317646 (US, A1)
特開2011-233119 (JP, A)
特開2007-181070 (JP, A)
米国特許出願公開第2015/0281774 (US, A1)
特開2010-191723 (JP, A)
特開2009-268036 (JP, A)
国際公開第2015/180652 (WO, A1)
特開2012-123727 (JP, A)
- (58)調査した分野 (Int.Cl., DB名)
H04N 21/00 - 21/858
H04N 17/00