

(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
2022年9月1日(01.09.2022)



(10) 国際公開番号

WO 2022/180861 A1

(51) 国際特許分類:
H04N 7/15 (2006.01)

上ハイツ 301 株式会社 I'mbesideyou 内 Tokyo (JP).

(21) 国際出願番号: PCT/JP2021/007578

(22) 国際出願日: 2021年2月26日(26.02.2021)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 株式会社 I'mbesideyou (I'MBESIDEYOU INC.) [JP/JP]; 〒1570066
東京都世田谷区成城 2 丁目 3 4 番 2 号 村
上ハイツ 301 Tokyo (JP).

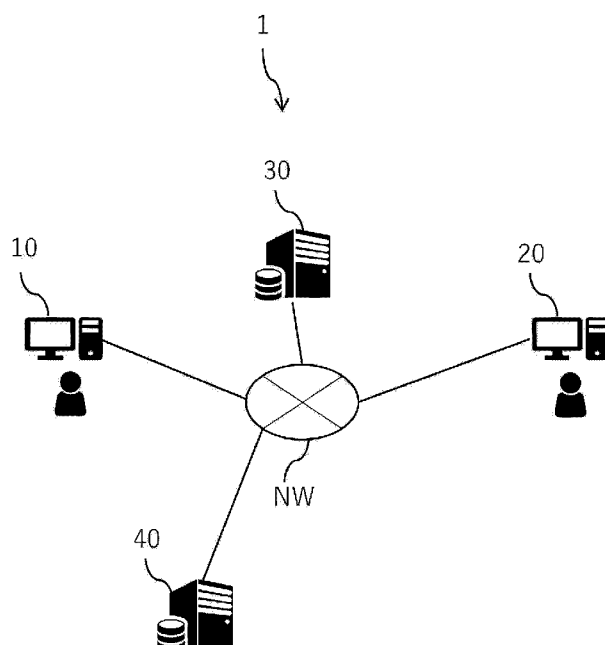
(72) 発明者: 神谷 渉三(KAMIYA Shozo); 〒1570066
東京都世田谷区成城 2 丁目 3 4 番 2 号 村

(74) 代理人: 中 畑 稔 (NAKAHATA MINORU);
〒1040028 東京都中央区八重洲 2 丁目 8 番 8
号 大星八重洲ビル 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, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT,

(54) Title: VIDEO SESSION EVALUATION TERMINAL, VIDEO SESSION EVALUATION SYSTEM, AND VIDEO SESSION EVALUATION PROGRAM

(54) 発明の名称: ビデオセッション評価端末、ビデオセッション評価システム及びビデオセッション評価プログラム



(57) Abstract: [Problem] To evaluate an online session by evaluating video acquired in the online session. [Solution] A system according to the present disclosure comprises: an acquisition means that acquires at least a video; a face recognition means that recognizes, for each predetermined frame, at least a face image of a person of interest included in the video; a voice recognition means that recognizes at least a voice of the person of interest included in the video; an intonation acquisition means that extracts intonation information of the recognized voice; and an evaluation means that calculates,

[続葉有]

QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, 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:

- 国際調査報告 (条約第21条(3))

on the basis of both the recognized face image and the intonation information, an evaluation value according to a specific viewpoint.

(57) 要約: 【課題】オンラインセッションにおいて取得された動画像を評価することにより、オンラインセッション自体の評価を行うこと。【解決手段】本開示のシステムは少なくとも動画像を取得する取得手段と、前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、認識した前記音声の抑揚情報を抽出する抑揚取得手段と、認識した前記顔画像及び前記抑揚情報の双方に基づいて所定の観点による評価値を算出する評価手段と、を備える。

明 細 書

発明の名称：

ビデオセッション評価端末、ビデオセッション評価システム及びビデオセッション評価プログラム

技術分野

[0001] 本開示は、ビデオセッション評価端末、ビデオセッション評価システム及びビデオセッション評価プログラムに関する。

背景技術

[0002] 従来、発言者の発言に対して他者が受ける感情を解析する技術が知られている（例えば、特許文献1参照）。また、対象者の表情の変化を長期間にわたり時系列的に解析し、その間に抱いた感情を推定する技術も知られている（例えば、特許文献2参照）。さらに、感情の変化に最も影響を与えた要素を特定する技術も知られている（例えば、特許文献3～5参照）。さらにまた、対象者の普段の表情と現在の表情とを比較して、表情が暗い場合にアラートを発する技術も知られている（例えば、特許文献6参照）。また、対象者の平常時（無表情時）の表情と現在の表情とを比較して、対象者の感情の度合いを判定するようにした技術も知られている（例えば、特許文献7～9参照）。更に、また、組織としての感情や、個人が感じるグループ内の雰囲気进行分析する技術も知られている（例えば、特許文献10、11参照）。

先行技術文献

特許文献

- [0003] 特許文献1：特開2019-58625号公報
特許文献2：特開2016-149063号公報
特許文献3：特開2020-86559号公報
特許文献4：特開2000-76421号公報
特許文献5：特開2017-201499号公報
特許文献6：特開2018-112831号公報

特許文献7：特開2011-154665号公報

特許文献8：特開2012-8949号公報

特許文献9：特開2013-300号公報

特許文献10：特開2011-186521号公報

特許文献11：WO15/174426号公報

発明の概要

発明が解決しようとする課題

[0004] 上述したすべての技術は、現実空間におけるコミュニケーションが主である状況におけるサブ的な機能にすぎない。即ち、昨今の業務のDX（Digital Transformation）化や、世界的な感染症の流行等を受け、業務や授業等のコミュニケーションがオンラインで行われることが主とされる状況に生まれたものではない。

[0005] 本発明は、オンラインコミュニケーションが主となる状況において、より効率的なコミュニケーションを行うために、交わされたコミュニケーションを客観的に評価することを目的とする。

課題を解決するための手段

[0006] 本発明によれば、

第1ユーザと第2ユーザとのオンラインセッションに関する動画像を取得する取得手段と、

前記動画像内に含まれる前記第1ユーザ及び前記第2ユーザの少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて、複数の観点による評価値を算出する評価手段と、

前記評価値に基づいて、前記第2ユーザの前記第1ユーザに対するマッチ度を判定する判定手段とを備える、

動画像分析システムが得られる。

[0007] 本発明によれば、

対象者を撮影することによって得られる複数の動画像を取得する動画像取得部と、

前記動画像取得部により取得された動画像に基づいて、前記対象者について生体反応の変化を解析する生体反応解析部と、

前記生体反応解析部により前記対象者について解析された前記生体反応の変化に基づいて、複数の前記動画像間で前記対象者について平準化された評価基準に従って前記対象者の感情の度合いを評価する感情評価部とを備える動画像分析システムが得られる。

[0008] 本発明によれば、

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて所定の感情情報に分類する評価手段と、

分類された前記感情情報へのアノテーション操作を受け付けるアノテーション受付手段と、を備える、
動画像分析システムが得られる。

[0009] 本発明によれば、

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて複数の観点による評価値を算出する評価手段と、

複数の前記観点による前記評価値それぞれについて所定期間における平均値を当該対象者へ提供する評価値提供手段と、を備える、
動画像分析システムが得られる。

[0010] 本発明によれば、

第1ユーザと第2ユーザとのオンラインセッションに関する動画像を取得する取得手段と、

前記動画像内に含まれる前記第1ユーザの少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記第2ユーザの少なくとも音声を認識する音声認識手段と、

少なくとも認識した前記顔画像に基づいて所定の観点による評価値を算出する評価手段と、

少なくとも認識した前記音声内の所定のキーワードを検出するキーワード検出手段と、

前記キーワードを検出したときにおける前記評価値に基づいて、所定のアラートを送信するアラート送信手段とを備える、

動画像分析システムが得られる。

[0011] 本発明によれば、

少なくとも対象者が映っている複数の動画像を取得する取得手段と、

前記動画像のうち評価対象動画像に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記音声に含まれる単語のうち、前記評価対象動画像以外の前記動画像には含まれていなかった単語を抽出する固有単語抽出手段と、

抽出した前記単語をその発言頻度に応じたサイズに変換してテキスト表示するテキスト表示手段とを備える、

動画像分析システムが得られる。

[0012] 本発明によれば、

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて所定の観点による評価値を算出する評価手段と、

認識した前記音声に含まれる単語をテキストに変換して表示するテキスト変換手段と、

変換した前記テキストのサイズをその発言頻度に応じた大きさに設定するサイズ設定手段と、

変換した前記テキストの色を前記評価値に応じた色に設定する色設定手段と、

を備える、

動画像分析システムが得られる。

[0013] 本発明によれば、

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて疲労度を評価する評価手段とを備える、

動画像分析システムが得られる。

[0014] 本発明によれば、

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる発言者毎に発言を認識する音声認識手段と、

前記発言と前記対象者とを関連付けて時系列に並べて表示するオブジェクトを生成するオブジェクト生成部と、を備える

動画像分析システムが得られる。

[0015] 本発明によれば、
少なくとも動画像を取得する取得手段と、
前記動画像内に含まれる対象者毎に発言を認識する音声認識手段と、
前記発言に対応する発言オブジェクトを前記対象者と関連付けてプロットする発言オブジェクト生成部と、を備える
動画像分析システムが得られる。

[0016] 本発明によれば、
少なくとも動画像を取得する取得手段と、
前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、
前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、
認識した前記音声の抑揚情報を抽出する抑揚取得手段と、
認識した前記顔画像及び前記抑揚情報の双方に基づいて所定の観点による評価値を算出する評価手段と、を備える
動画像分析システムが得られる。

[0017] 本発明によれば、
少なくとも動画像を取得する取得手段と、
前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、
前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、
前記動画のコンテキスト情報を取得するコンテキスト取得手段と、
認識した前記顔画像及び前記音声の双方に基づいて所定の観点による評価値を算出する評価手段と、
前記コンテキスト情報を用いて当該評価値を補正する補正手段と、を備える、

動画像分析システム。

発明の効果

[0018] 本開示によれば、ビデオセッションの動画像を分析評価することにより、特に内容に関する評価を客観的に行うことができる。

[0019] 特に、本発明によれば、オンラインコミュニケーションが主となる状況において、より効率的なコミュニケーションを行うために、交わされたコミュニケーションを客観的に評価することができる。

図面の簡単な説明

[0020] [図1]本発明の実施の形態によるシステム全体図を示す図である。

[図2]本発明の実施の形態による評価端末の機能ブロック図の一例である。

[図3]本発明の実施の形態による評価端末の機能構成例1を示す図である。

[図4]本発明の実施の形態による評価端末の機能構成例2を示す図である。

[図5]本発明の実施の形態による評価端末の機能構成例3を示す図である。

[図6]図6の機能構成例3による画面表示例である。

[図7]図6の機能構成例3による他の画面表示例である。

[図8]本発明の実施の形態による評価端末の機能構成例3の他の構成を示す図である。

[図9]本発明の実施の形態による評価端末の機能構成例3の他の構成を示す図である。

[図10]本発明の第1の実施の形態によるシステムを示す図である。

[図11]本発明の第1の実施の形態によるシステムを示す図である。

[図12]本発明の第2の実施の形態によるシステムを示す図である。

[図13]本発明の第2の実施の形態によるシステムを示す図である。

[図14]本発明の第3の実施の形態によるシステムを示す図である。

[図15]本発明の第4の実施の形態によるシステムを示す図である。

[図16]本発明の第5の実施の形態によるシステムを示す図である。

[図17]本発明の第6の実施の形態によるシステムを示す図である。

[図18]本発明の第6の実施の形態によるシステムを示す図である。

[図19]本発明の第7の実施の形態によるシステムを示す図である。

[図20]本発明の第7の実施の形態によるシステムを示す図である。

[図21]本発明の第8の実施の形態によるシステムを示す図である。

[図22]本発明の第9の実施の形態によるシステムを示す図である。

[図23]本発明の第10の実施の形態によるシステムを示す図である。

[図24]本発明の第10の実施の形態によるシステムを示す図である。

[図25]本発明の第11の実施の形態によるシステムを示す図である。

発明を実施するための形態

[0021] 本開示の実施形態の内容を列記して説明する。本開示は、以下のような構成を備える。

[項目1]

第1ユーザと第2ユーザとのオンラインセッションに関する動画像を取得する取得手段と、

前記動画像内に含まれる前記第1ユーザ及び前記第2ユーザの少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて、複数の観点による評価値を算出する評価手段と、

前記評価値に基づいて、前記第2ユーザの前記第1ユーザに対するマッチ度を判定する判定手段とを備える、

動画像分析システム。

[項目2]

対象者を撮影することによって得られる複数の動画像を取得する動画像取得部と、

前記動画像取得部により取得された動画像に基づいて、前記対象者について生体反応の変化を解析する生体反応解析部と、

前記生体反応解析部により前記対象者について解析された前記生体反応の

変化に基づいて、複数の前記動画像間で前記対象者について平準化された評価基準に従って前記対象者の感情の度合いを評価する感情評価部とを備える動画像分析システム。

[項目 3]

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて所定の感情情報に分類する評価手段と、

分類された前記感情情報へのアノテーション操作を受け付けるアノテーション受付手段と、を備える、
動画像分析システム。

[項目 4]

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて複数の観点による評価値を算出する評価手段と、

複数の前記観点による前記評価値それぞれについて所定期間における平均値を当該対象者へ提供する評価値提供手段と、を備える、
動画像分析システム。

[項目 5]

第 1 ユーザと第 2 ユーザとのオンラインセッションに関する動画像を取得する取得手段と、

前記動画像内に含まれる前記第1ユーザの少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記第2ユーザの少なくとも音声を認識する音声認識手段と、

少なくとも認識した前記顔画像に基づいて所定の観点による評価値を算出する評価手段と、

少なくとも認識した前記音声内の所定のキーワードを検出するキーワード検出手段と、

前記キーワードを検出したときにおける前記評価値に基づいて、所定のアラートを送信するアラート送信手段とを備える、

動画像分析システム。

[項目6]

少なくとも対象者が映っている複数の動画像を取得する取得手段と、

前記動画像のうち評価対象動画像に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記音声に含まれる単語のうち、前記評価対象動画像以外の前記動画像には含まれていなかった単語を抽出する固有単語抽出手段と、

抽出した前記単語をその発言頻度に応じたサイズに変換してテキスト表示するテキスト表示手段とを備える、

動画像分析システム。

[項目7]

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて所定の観点による評価値を算出する評価手段と、

認識した前記音声に含まれる単語をテキストに変換して表示するテキスト変換手段と、

変換した前記テキストのサイズをその発言頻度に応じた大きさに設定するサイズ設定手段と、

変換した前記テキストの色を前記評価値に応じた色に設定する色設定手段と、

を備える、

動画像分析システム。

[項目 8]

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記顔画像及び前記音声の双方に基づいて疲労度を評価する評価手段とを備える、

動画像分析システム。

[項目 9]

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる発言者毎に発言を認識する音声認識手段と、

前記発言と前記対象者とを関連付けて時系列に並べて表示するオブジェクトを生成するオブジェクト生成部と、を備える

動画像分析システム。

[項目 10]

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者毎に発言を認識する音声認識手段と、

前記発言に対応する発言オブジェクトを前記対象者と関連付けてプロットする発言オブジェクト生成部と、を備える

動画像分析システム。

[項目 1 1]

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

認識した前記音声の抑揚情報を抽出する抑揚取得手段と、

認識した前記顔画像及び前記抑揚情報の双方に基づいて所定の観点による評価値を算出する評価手段と、を備える

動画像分析システム。

[項目 1 2]

少なくとも動画像を取得する取得手段と、

前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、

前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、

前記動画のコンテキスト情報を取得するコンテキスト取得手段と、

認識した前記顔画像及び前記音声の双方に基づいて所定の観点による評価値を算出する評価手段と、

前記コンテキスト情報を用いて当該評価値を補正する補正手段と、を備える、

動画像分析システム。

[0022] 以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

[0023] <基本機能>

本実施形態のビデオセッション評価システムは、複数人でビデオセッション（以下、一方向及び双方向含めてオンラインセッションという）が行われる環境において、当該複数人の中の解析対象者について他者とは異なる特異的な感情（自分または他人の言動に対して起こる気持ち。快・不快またはその程度など）を解析し評価するシステムである。オンラインセッションは、例えばオンライン会議、オンライン授業、オンラインチャットなどであり、複数の場所に設置された端末をインターネットなどの通信ネットワークを介してサーバに接続し、当該サーバを通じて複数の端末間で動画像をやり取りできるようにしたものである。オンラインセッションで扱う動画像には、端末を使用するユーザの顔画像や音声が含まれる。また、動画像には、複数のユーザが共有して閲覧する資料などの画像も含まれる。各端末の画面上に顔画像と資料画像とを切り替えて何れか一方のみを表示させたり、表示領域を分けて顔画像と資料画像とを同時に表示させたりすることが可能である。また、複数人のうち1人の画像を全画面表示させたり、一部または全部のユーザの画像を小画面に分割して表示させたりすることが可能である。端末を使用してオンラインセッションに参加する複数のユーザのうち、何れか1人または複数人を解析対象者として指定することが可能である。例えば、オンラインセッションの主導者、進行者または管理者（以下、まとめて主催者という）が何れかのユーザを解析対象者として指定する。オンラインセッションの主催者は、例えばオンライン授業の講師、オンライン会議の議長やファシリテータ、コーチングを目的としたセッションのコーチなどである。オンラインセッションの主催者は、オンラインセッションに参加する複数のユーザの中の一人であるのが普通であるが、オンラインセッションに参加しない別人であってもよい。なお、解析対象者を指定せず全ての参加者を解析対象としてもよい。また、オンラインセッションの主導者、進行者または管理者（以下、まとめて主催者という）が何れかのユーザを解析対象者として指定することも可能である。オンラインセッションの主催者は、例えばオンライン授業の講師、オンライン会議の議長やファシリテータ、コーチングを目的と

したセッションのコーチなどである。オンラインセッションの主催者は、オンラインセッションに参加する複数のユーザの中の一人であるのが普通であるが、オンラインセッションに参加しない別人であってもよい。

[0024] 本実施の形態によるビデオセッション評価システムは、複数の端末間においてビデオセッションセッションが確立された場合に、当該ビデオセッションから取得される少なくとも動画像を表示される。表示された動画像は、端末によって取得され、動画像内に含まれる少なくとも顔画像を所定のフレーム単位ごとに識別される。その後、識別された顔画像に関する評価値が算出される。当該評価値は必要に応じて共有される。特に、本実施の形態においては、取得した動画像は当該端末に保存され、端末上で分析評価され、その結果が当該端末のユーザに提供される。従って、例えば個人情報を含むビデオセッションや機密情報を含むビデオセッションであっても、その動画自体を外部の評価機関等に提供することなく分析評価できる。また、必要に応じて、当該評価結果（評価値）だけを外部端末に提供することによって、結果を可視化したり、クロス分析等行うことができる。

[0025] 図1に示されるように、本実施の形態によるビデオセッション評価システムは、少なくともカメラ部及びマイク部等の入力部と、ディスプレイ等の表示部とスピーカー等の出力部とを有するユーザ端末10、20と、ユーザ端末10、20に双方向のビデオセッションを提供するビデオセッションサービス端末30と、ビデオセッションに関する評価の一部を行う評価端末40とを備えている。

[0026] <ハードウェア構成例>

以下に説明する各機能ブロック、機能単位、機能モジュールは、例えばコンピュータに備えられたハードウェア、DSP（Digital Signal Processor）、ソフトウェアの何れによっても構成することが可能である。例えばソフトウェアによって構成する場合、実際にはコンピュータのCPU、RAM、ROMなどを備えて構成され、RAMやROM、ハードディスクまたは半導体メモリ等の記録媒体に記憶されたプログラムが動作することによって実現

される。本明細書において説明するシステム及び端末による一連の処理は、ソフトウェア、ハードウェア、及びソフトウェアとハードウェアとの組合せのいずれを用いて実現されてもよい。本実施形態に係る情報共有支援装置 10 の各機能を実現するためのコンピュータプログラムを作製し、PC 等に実装することが可能である。また、このようなコンピュータプログラムが格納された、コンピュータで読み取り可能な記録媒体も提供することが可能である。記録媒体は、例えば、磁気ディスク、光ディスク、光磁気ディスク、フラッシュメモリ等である。また、上記のコンピュータプログラムは、記録媒体を用いずに、例えばネットワークを介して配信されてもよい。

[0027] 本実施の形態による評価端末は、ビデオセッションサービス端末から動画像を取得し、当該動画像内に含まれる少なくとも顔画像を所定のフレーム単位ごとに識別すると共に、顔画像に関する評価値を算出する（詳しくは後述する）。

＜動画の取得方法＞

図 3 に示されるように、ビデオセッションサービス端末が提供するビデオセッションサービス（以下、単に「本サービス」と言うことがある）は、ユーザ端末 10、20 に対して双方向に画像および音声によって通信が可能となるものである。本サービスは、ユーザ端末のディスプレイに相手のユーザ端末のカメラ部で取得した動画像を表示し、相手のユーザ端末のマイク部で取得した音声をスピーカーから出力可能となっている。また、本サービスは双方の又はいずれかのユーザ端末によって、動画像及び音声（これらを合わせて「動画像等」という）を少なくともいずれかのユーザ端末上の記憶部に記録（レコーディング）することが可能に構成されている。記録された動画像情報 Vs（以下「記録情報」という）は、記録を開始したユーザ端末にキャッシュされつついずれかのユーザ端末のローカルのみに記録されることとなる。ユーザは、必要があれば当該記録情報を本サービスの利用の範囲内で自分で視聴、他者に共有等行うこともできる。

[0028] ＜機能構成例 1＞

図4は、本実施形態による構成例を示すブロック図である。図4に示すように、本実施形態のビデオセッション評価システムは、ユーザ端末10が有する機能構成として実現される。すなわち、ユーザ端末10はその機能として、動画像取得部11、生体反応解析部12、特異判定部13、関連事象特定部14、クラスタリング部15および解析結果通知部16を備えている。

[0029] 動画像取得部11は、オンラインセッション中に各端末が備えるカメラにより複数人（複数のユーザ）を撮影することによって得られる動画像を各端末から取得する。各端末から取得する動画像は、各端末の画面上に表示されるように設定されているものか否かは問わない。すなわち、動画像取得部11は、各端末に表示中の動画像および非表示中の動画像を含めて、動画像を各端末から取得する。

[0030] 生体反応解析部12は、動画像取得部11により取得された動画像（画面上に表示中のものか否かは問わない）に基づいて、複数人のそれぞれについて生体反応の変化を解析する。本実施形態において生体反応解析部12は、動画像取得部11により取得された動画像を画像のセット（フレーム画像の集まり）と音声とに分離し、それぞれから生体反応の変化を解析する。

[0031] 例えば、生体反応解析部12は、動画像取得部11により取得された動画像から分離したフレーム画像を用いてユーザの顔画像を解析することにより、表情、目線、脈拍、顔の動きの少なくとも1つに関する生体反応の変化を解析する。また、生体反応解析部12は、動画像取得部11により取得された動画像から分離した音声を解析することにより、ユーザの発言内容、声質の少なくとも1つに関する生体反応の変化を解析する。

[0032] 人は感情が変化すると、それが表情、目線、脈拍、顔の動き、発言内容、声質などの生体反応の変化となって現れる。本実施形態では、ユーザの生体反応の変化を解析することを通じて、ユーザの感情の変化を解析する。本実施形態において解析する感情は、一例として、快／不快の程度である。本実施形態において生体反応解析部12は、生体反応の変化を所定の基準に従って数値化することにより、生体反応の変化の内容を反映させた生体反応指標

値を算出する。

[0033] 表情の変化の解析は、例えば以下のようにして行う。すなわち、フレーム画像ごとに、フレーム画像の中から顔の領域を特定し、事前に機械学習させた画像解析モデルに従って特定した顔の表情を複数に分類する。そして、その分類結果に基づいて、連続するフレーム画像間でポジティブな表情変化が起きているか、ネガティブな表情変化が起きているか、およびどの程度の大きさの表情変化が起きているかを解析し、その解析結果に応じた表情変化指標値を出力する。

[0034] 目線の変化の解析は、例えば以下のようにして行う。すなわち、フレーム画像ごとに、フレーム画像の中から目の領域を特定し、両目の向きを解析することにより、ユーザがどこを見ているかを解析する。例えば、表示中の話者の顔を見ているか、表示中の共有資料を見ているか、画面の外を見ているかななどを解析する。また、目線の動きが大きい小さいか、動きの頻度が多いか少ないかななどを解析するようにしてもよい。目線の変化はユーザの集中度にも関連する。生体反応解析部12は、目線の変化の解析結果に応じた目線変化指標値を出力する。

[0035] 脈拍の変化の解析は、例えば以下のようにして行う。すなわち、フレーム画像ごとに、フレーム画像の中から顔の領域を特定する。そして、顔の色情報（RGBのG）の数値を捉える学習済みの画像解析モデルを用いて、顔表面のG色の変化を解析する。その結果を時間軸に合わせて並べることによって色情報の変化を表した波形を形成し、この波形から脈拍を特定する。人は緊張すると脈拍が速くなり、気持ちが落ち着くと脈拍が遅くなる。生体反応解析部12は、脈拍の変化の解析結果に応じた脈拍変化指標値を出力する。

[0036] 顔の動きの変化の解析は、例えば以下のようにして行う。すなわち、フレーム画像ごとに、フレーム画像の中から顔の領域を特定し、顔の向きを解析することにより、ユーザがどこを見ているかを解析する。例えば、表示中の話者の顔を見ているか、表示中の共有資料を見ているか、画面の外を見ているかななどを解析する。また、顔の動きが大きい小さいか、動きの頻度が多

いか少ないかなどを解析するようにしてもよい。顔の動きと目線の動きとを合わせて解析するようにしてもよい。例えば、表示中の話者の顔をまっすぐ見ているか、上目遣いまたは下目使いに見ているか、斜めから見ているかなどを解析するようにしてもよい。生体反応解析部 12 は、顔の向きの変化の解析結果に応じた顔向き変化指標値を出力する。

[0037] 発言内容の解析は、例えば以下のようにして行う。すなわち、生体反応解析部 12 は、指定した時間（例えば、30～150 秒程度の時間）の音声について公知の音声認識処理を行うことによって音声を文字列に変換し、当該文字列を形態素解析することにより、助詞、冠詞などの会話を表す上で不要なワードを取り除く。そして、残ったワードをベクトル化し、ポジティブな感情変化が起きているか、ネガティブな感情変化が起きているか、およびどの程度の大きさの感情変化が起きているかを解析し、その解析結果に応じた発言内容指標値を出力する。

[0038] 声質の解析は、例えば以下のようにして行う。すなわち、生体反応解析部 12 は、指定した時間（例えば、30～150 秒程度の時間）の音声について公知の音声解析処理を行うことによって音声の音響的特徴を特定する。そして、その音響的特徴に基づいて、ポジティブな声質変化が起きているか、ネガティブな声質変化が起きているか、およびどの程度の大きさの声質変化が起きているかを解析し、その解析結果に応じた声質変化指標値を出力する。

[0039] 生体反応解析部 12 は、以上のようにして算出した表情変化指標値、目線変化指標値、脈拍変化指標値、顔向き変化指標値、発言内容指標値、声質変化指標値の少なくとも 1 つを用いて生体反応指標値を算出する。例えば、表情変化指標値、目線変化指標値、脈拍変化指標値、顔向き変化指標値、発言内容指標値および声質変化指標値を重み付け計算することにより、生体反応指標値を算出する。

[0040] 特異判定部 13 は、解析対象者について解析された生体反応の変化が、解析対象者以外の他者について解析された生体反応の変化と比べて特異的か否

かを判定する。本実施形態において、特異判定部 1 3 は、生体反応解析部 1 2 により複数のユーザのそれぞれについて算出された生体反応指標値に基づいて、解析対象者について解析された生体反応の変化が他者と比べて特異的か否かを判定する。

[0041] 例えば、特異判定部 1 3 は、生体反応解析部 1 2 により複数人のそれぞれについて算出された生体反応指標値の分散を算出し、解析対象者について算出された生体反応指標値と分散との対比により、解析対象者について解析された生体反応の変化が他者と比べて特異的か否かを判定する。

[0042] 解析対象者について解析された生体反応の変化が他者と比べて特異的である場合として、次の 3 パターンが考えられる。1 つ目は、他者については特に大きな生体反応の変化が起きていないが、解析対象者について比較的大きな生体反応の変化が起きた場合である。2 つ目は、解析対象者については特に大きな生体反応の変化が起きていないが、他者について比較的大きな生体反応の変化が起きた場合である。3 つ目は、解析対象者についても他者についても比較的大きな生体反応の変化が起きているが、変化の内容が解析対象者と他者とで異なる場合である。

[0043] 関連事象特定部 1 4 は、特異判定部 1 3 により特異的であると判定された生体反応の変化が起きたときに解析対象者、他者および環境の少なくとも 1 つに関して発生している事象を特定する。例えば、関連事象特定部 1 4 は、解析対象者について特異的な生体反応の変化が起きたときにおける解析対象者自身の言動を動画像から特定する。また、関連事象特定部 1 4 は、解析対象者について特異的な生体反応の変化が起きたときにおける他者の言動を動画像から特定する。また、関連事象特定部 1 4 は、解析対象者について特異的な生体反応の変化が起きたときにおける環境を動画像から特定する。環境は、例えば画面に表示中の共有資料、解析対象者の背景に写っているものなどである。

[0044] クラスタリング部 1 5 は、特異判定部 1 3 により特異的であると判定された生体反応の変化（例えば、目線、脈拍、顔の動き、発言内容、声質のうち

1つまたは複数の組み合わせ）と、当該特異的な生体反応の変化が起きたときに発生している事象（関連事象特定部14により特定された事象）との相関の程度を解析し、相関が一定レベル以上であると判定された場合に、その相関の解析結果に基づいて解析対象者または事象をクラスタリングする。

[0045] 例えば、特異的な生体反応の変化がネガティブな感情変化に相当するものであり、当該特異的な生体反応の変化が起きたときに発生している事象もネガティブな事象である場合には一定レベル以上の相関が検出される。クラスタリング部15は、その事象の内容やネガティブな度合い、相関の大きさなどに応じて、あらかじめセグメント化した複数の分類の何れかに解析対象者または事象をクラスタリングする。

[0046] 同様に、特異的な生体反応の変化がポジティブな感情変化に相当するものであり、当該特異的な生体反応の変化が起きたときに発生している事象もポジティブな事象である場合には一定レベル以上の相関が検出される。クラスタリング部15は、その事象の内容やポジティブな度合い、相関の大きさなどに応じて、あらかじめセグメント化した複数の分類の何れかに解析対象者または事象をクラスタリングする。

[0047] 解析結果通知部16は、特異判定部13により特異的であると判定された生体反応の変化、関連事象特定部14により特定された事象、およびクラスタリング部15によりクラスタリングされた分類の少なくとも1つを、解析対象者の指定者（解析対象者またはオンラインセッションの主催者）に通知する。

[0048] 例えば、解析結果通知部16は、解析対象者について他者とは異なる特異的な生体反応の変化が起きたとき（上述した3パターンの何れか。以下同様）に発生している事象として解析対象者自身の言動を解析対象者自身に通知する。これにより、解析対象者は、自分がある言動を行ったときに他者とは違う感情を持っていることを把握することができる。このとき、解析対象者について特定された特異的な生体反応の変化も併せて解析対象者に通知するようにしてもよい。さらに、対比される他者の生体反応の変化を更に解析対

象者に通知するようにしてもよい。

[0049] 例えば、解析対象者が普段どおりの感情で特に意識せずに行った言動、または、解析対象者がある感情を伴って特に意識して行った言動に対して他者が受けた感情と、言動の際に解析対象者自身が抱いていた感情とが相違している場合に、そのときの解析対象者自身の言動が解析対象者に通知される。これにより、自分の意識に反して他者の受けが良い言動や他者の受けが良くない言動などを発見することも可能である。

[0050] また、解析結果通知部 16 は、解析対象者について他者とは異なる特異的な生体反応の変化が起きたときに発生している事象を、特異的な生体反応の変化と共にオンラインセッションの主催者に通知する。これにより、オンラインセッションの主催者は、指定した解析対象者に特有の現象として、どのような事象がどのような感情の変化に影響を与えているのかを知ることができる。そして、その把握した内容に応じて適切な処置を解析対象者に対して行うことが可能となる。

[0051] また、解析結果通知部 16 は、解析対象者について他者とは異なる特異的な生体反応の変化が起きたときに発生している事象または解析対象者のクラスタリング結果をオンラインセッションの主催者に通知する。これにより、オンラインセッションの主催者は、指定した解析対象者がどの分類にクラスタリングされたかによって、解析対象者に特有の行動の傾向を把握したり、今後起こり得る行動や状態などを予測したりすることができる。そして、それに対して適切な処置を解析対象者に対して行うことが可能となる。

[0052] なお、上記実施形態では、生体反応の変化を所定の基準に従って数値化することによって生体反応指標値を算出し、複数人のそれぞれについて算出された生体反応指標値に基づいて、解析対象者について解析された生体反応の変化が他者と比べて特異的な否かを判定する例について説明したが、この例に限定されない。例えば、以下のようにしてもよい。

[0053] すなわち、生体反応解析部 12 は、複数人のそれぞれについて目線の動きを解析して目線の方向を示すヒートマップを生成する。特異判定部 13 は、

生体反応解析部 12 により解析対象者について生成されたヒートマップと他者について生成されたヒートマップとの対比により、解析対象者について解析された生体反応の変化が、他者について解析された生体反応の変化と比べて特異的か否かを判定する。

[0054] このように、本実施の形態においては、ビデオセッションの動画像をユーザ端末 10 のローカルストレージに保存し、ユーザ端末 10 上で上述した分析を行うこととしている。ユーザ端末 10 のマシンスペックに依存する可能性があるとはいえ、動画像の情報を外部に提供することなく分析することが可能となる。

[0055] <機能構成例 2>

図 5 に示すように、本実施形態のビデオセッション評価システムは、機能構成として、動画像取得部 11、生体反応解析部 12 および反応情報提示部 13a を備えていてもよい。

[0056] 反応情報提示部 13a は、画面に表示されていない参加者を含めて生体反応解析部 12a により解析された生体反応の変化を示す情報を提示する。例えば、反応情報提示部 13a は、生体反応の変化を示す情報をオンラインセッションの主導者、進行者または管理者（以下、まとめて主催者という）に提示する。オンラインセッションの主催者は、例えばオンライン授業の講師、オンライン会議の議長やファシリテータ、コーチングを目的としたセッションのコーチなどである。オンラインセッションの主催者は、オンラインセッションに参加する複数のユーザの中の一人であるのが普通であるが、オンラインセッションに参加しない別人であってもよい。

[0057] このようにすることにより、オンラインセッションの主催者は、複数人でオンラインセッションが行われる環境において、画面に表示されていない参加者の様子も把握することができる。

[0058] <機能構成例 3>

図 6 は、本実施形態による構成例を示すブロック図である。図 6 に示すように、本実施形態のビデオセッション評価システムは、機能構成として、上

述した実施の形態１と類似する機能については同一つの参照符号を付して説明を省略することがある。

[0059] 本実施の形態によるシステムは、ビデオセッションの映像を取得するカメラ部及び音声を取得するマイク部と、動画像を分析及び評価する解析部、取得した動画像を評価することによって得られた情報に基づいて表示オブジェクト（後述する）を生成するオブジェクト生成部、前記ビデオセッション実行中にビデオセッションの動画像と表示オブジェクトの両方を表示する表示部と、を備えている。

[0060] 解析部は、上述した説明と同様に、動画像取得部１１、生体反応解析部１２、特異判定部１３、関連事象特定部１４、クラスタリング部１５および解析結果通知部１６を備えている。各要素の機能については上述したとおりである。

[0061] 図７に示されるように、オブジェクト生成部は、解析部によってビデオセッションから取得される動画像を解析した結果に基づいて、必要に応じて、当該認識した顔の部分を示すオブジェクト５０と、上述した分析・評価した内容を示す情報１００を当該動画像に重畳して表示する。当該オブジェクト５０は、複数人の顔が動画像内に移っている場合には、複数人全員の顔を識別し、表示することとしてもよい。

[0062] また、オブジェクト５０は、例えば、相手側の端末において、ビデオセッションのカメラ機能を停止している場合（即ち、物理的にカメラを覆う等ではなく、ビデオセッションのアプリケーション内においてソフトウェア的に停止している場合）であっても、相手側のカメラで相手の顔を認識していた場合には、相手の顔が位置している部分にオブジェクト５０やオブジェクト１００を表示することとしてもよい。これにより、カメラ機能がオフになっていたとしても、相手側が端末の前にいることがお互い確認することが可能となる。この場合、例えば、ビデオセッションのアプリケーションにおいては、カメラから取得した情報を非表示にする一方、解析部によって認識された顔に対応するオブジェクト５０やオブジェクト１００のみを表示すること

としてもよい。また、ビデオセッションから取得される映像情報と、解析部によって認識され得られた情報とを異なる表示レイヤーに分け、前者の情報に関するレイヤーを非表示にすることとしてもよい。

[0063] オブジェクト50やオブジェクト100は、複数の動画像を表示する領域がある場合には、すべての領域又は一部の領域のみに表示することとしてもよい。例えば、図8に示されるように、ゲスト側の動画像のみに表示することとしてもよい。

[0064] 以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

[0065] 本明細書において説明した装置は、単独の装置として実現されてもよく、一部または全部がネットワークで接続された複数の装置（例えばクラウドサーバ）等により実現されてもよい。例えば、各端末10の制御部110およびストレージ130は、互いにネットワークで接続された異なるサーバにより実現されてもよい。

[0066] 即ち、本システムは、ユーザ端末10、20と、ユーザ端末10、20に双方向のビデオセッションを提供するビデオセッションサービス端末30と、ビデオセッションに関する評価を行う評価端末40とを含んでいるところ、以下のような構成のバリエーション組み合わせが考えられる。

（１）すべてをユーザ端末のみで処理

図9に示されるように、解析部による処理をビデオセッションを行っている端末で行うことにより、（一定の処理能力は必要なものの）ビデオセッションを行っている時間と同時に（リアルタイムに）分析・評価結果を得ることができる。

（２）ユーザ端末と評価端末とで処理

図10に示されるように、ネットワーク等で接続された評価端末に解析部を備えさせることとしてもよい。この場合、ユーザ端末で取得された動画像は、ビデオセッションと同時に又は事後的に評価端末に共有され、評価端末における解析部によって分析・評価されたのちに、オブジェクト50及びオブジェクト100の情報がユーザ端末に動画像データと共に又は別に（即ち、少なくとも解析データを含む情報が）共有され表示部に表示される。

[0067] 以上説明した機能構成例1乃至機能構成例3の各構成及びそれらの組み合わせを用いて、図10乃至図25に示す第1乃至第11の実施の形態によるシステムが実現する。

[0068] <第1の実施の形態>

図10及び図11を参照して、本発明の第1の実施の形態を説明する。本実施の形態によるシステムは、概略、人と人とのマッチング度合いを評価する。例えば、相手の反応を分析したり、相手別の特異性（普段は出ない表情等）を評価したり、自分の過去との特異性（同じ相手でも過去には出なかった表情等）、ニュートラルな状態との比較を行うことにより評価する。特に、当該マッチングは、講師と受講生が行うオンラインセッションに対して有効である。様々なタイプの講師と様々なタイプの受講生とのマッチングは当該講座を継続するうえでも重要である。

[0069] 図10に示されるように、本実施の形態によるシステムは、上述した解析部による評価結果から各人のタイプを決定するタイプ決定部と、マッチング度合いを判定するマッチング判定部とを備えている。

[0070] タイプ決定部は、評価結果と、タイプとが予め関連付けられたタイプデータベース（タイプDB）を参照することにより、ユーザ毎のタイプを決定（推定）する。マッチング判定部は、タイプ毎のマッチング度合いが予め定義されたマッチングデータベース（マッチングDB）を参照して、上記決定されたタイプ同士を事前に定義された相性度を利用してマッチング度合いを数値化する。マッチングDBの構築は、会話を引き出すのが得意なタイプの講師と、発言が苦手な受講生とのマッチング度合いを事前に定義しておくこと

などが例示できる。

[0071] なお、評価結果を取得した後に、2者間の会話と当該2者間の当該会話に対する評価とを含む教師データを学習した学習モデルを用いてマッチ度合いを判定することとしてもよい。この場合、実際にマッチングされた者同士の講義の結果をフィードバック（講師が自分に合っていた、合っていなかった、等）することとしてもよい。

[0072] 更に、図11に示されるように、講座開講の前に、受講生のタイプを事前に判定すべく、タイプ判定用のオンラインセッションが行われてもよい。システムは、タイプ判定のために行われたオンラインセッションの動画像を取得し（ステップS1101）、タイプの決定を行う（ステップS1102）。続いて、決定された受講生のタイプと講師のタイプとを一時的にマッチングを行う（ステップS1103）。受講生側の要望として、例えば「優しい先生がいい」「テンポの良い先生がいい」等のように講師に求める条件を事前にアンケート等で取得しておき、当該アンケートの結果から望ましいタイプを特定しておくこととしてもよい。システムは、このような情報を条件情報として取得する（ステップS1104）。システムは、条件情報を考慮して、一次マッチ度を補正して（ステップS1105）、補正後のマッチ度を提供する。これにより、システム側としては、厳しい講師がマッチング相手として適切と判断した場合であっても、受講生に「優しい性格が希望」という条件があった場合には、一次マッチ度が算出された複数の講師のマッチ度がそれぞれ補正されて「厳しくも優しい講師」が最適な講師として選定される。

[0073] <第2の実施の形態>

図12及び図13を参照して、本発明の第2の実施の形態を説明する。本実施の形態によるシステムは、概略、ユーザから得られた感情（評価値）に対して、そもそもその感情が出やすい人かどうか（例えば、もともと笑顔が多い人のhappyスコアは高くなりがち）というベースの感情との比較を考慮したり、感情が出たときの変位の大きさ（リアクションが小さい人と大

きい人とでは感情表出度合いが異なる)を評価したりして、正確に評価を行うものである。

[0074] 図12の(a)乃至(c)のグラフは、あるユーザの(a)生データ(時系列の感情スコア)、(b)生データを1分間隔のフレーム幅で取得した標準偏差(標準偏差処理)、(c)標準偏差を標準化(平均0、分散1の z score化)した(標準化処理)ものである。

[0075] システムは、(b)の評価値に基づいて、ユーザ毎の顔つきや定常時の表情の違いを考慮した評価を行う。例えば、定常時に笑顔であることが多いユーザは必然的に笑顔のスコア(happyスコア)が高くなってしまい、といった問題を改善することが可能となる。また、システムは(c)の評価に基づいて、ユーザ毎に感情表現の豊さの違いを考慮した評価を行う。例えば小さく笑う人と、大きく笑う人との差による問題を改善することが可能となる。

[0076] 図13の模式図に基づいて更に詳しく説明する。図13の(a)及び(b)は、それぞれ、ユーザaとユーザbのHappyスコア(幸福度を表す)グラフである。ユーザaとユーザbのそれぞれのスコアの平均を比べると、ユーザAの方が平均が高いことがわかる。即ち、ユーザAはユーザBに比べると笑顔が多い人であり、Happyスコアが必然的に高くなる傾向になることがわかる。また、それぞれの感情の幅(ST_A及びST_B)を比較すると、ユーザAの方がユーザBよりも感情表出の幅(即ち、リアクションの大きさ)が大きいことがわかる。

[0077] 上述した標準偏差処理と、標準化処理を行うことにより、その感情の表出頻度や、感情の表出の程度などの個人差を排除した数値の評価を行うことが可能となる。また、上述した解析部(例えば、図3等)に対して、標準偏差処理と及び標準化処理を施した教師データを利用した機械学習を行わせることにより、適切な学習を行うことが可能となる。即ち、本システムは、様々な動画像から得られた評価結果に対して標準偏差処理と、標準化処理を行うことにより教師データを生成するシステム(装置)としてもその機能を発揮

できる。

[0078] <第3の実施の形態>

図14を参照して、本発明の第3の実施の形態を説明する。本実施の形態によるシステムは、概略、評価結果に対して本人にアノテーション（ラベリング）させるものである。以下、講師（第1ユーザ）と受講生（第2ユーザ）のオンラインセッションを例に説明する。

[0079] システムは、上述したオンラインセッションの動画像を分析・評価し、受講生（第2ユーザ）その評価結果をグラフとして出力することにより可視化する。講師は、当該グラフに対してその感情の有無、その時の補足情報（状況、言動、行動、相手の行動情報等）を追加することができる。図示されるように、例えば、ユーザは、Happyスコアが高かった地点（Lab__1）に対して、その時の状況（「授業が盛り上がっていた」、「相手の反応がよかった」等の状況の情報）を関連付けることができる。また、Happyスコアが低かった地点（Lab__2）に対して、その時の状況（「何らかの課題をさせていた」、「厳しいことを伝えた」等の状況の情報）を関連付けることができる。これにより、何らかの課題に集中させていたからスコアが低かったのか、厳しいことを伝えたからスコアが低かったのか、改善方針に役立てることができる。このように、相手（第2ユーザ）の反応に対する、自分（第1ユーザ）のアノテーションを受け付けることにより、そのコミュニケーションが改善すべきものなのか、そうではないものなのかを判定することが可能となる。

[0080] なお、図示されるように、アノテーションは区間（Lab__3及びLab__4）に対して行わせることとしてもよい。この場合、「難しい単元を教えていた時間」「授業終盤のまとめの時間」のようなアノテーションも可能となる。

[0081] また、グラフ（プロットとされた評価値）と、アノテーションとのセットから教示データセットを生成し、解析部に機械学習させることとしてもよい。

[0082] <第4の実施の形態>

図15を参照して、本発明の第4の実施の形態を説明する。本実施の形態によるシステムは、概略、講師と受講生とで行われる講義のオンラインセッションにおいて、受講生の感情カルテとなり得るカルテ情報を生成して講師へ共有するものである。カルテの内容としては、例えば、受講生の感情毎の平均値、特徴、口癖、頻出の表情や、レーダーチャートなどによる感情表出のバランスと程度、落ち込んだときにかけられていた言葉のランキング、笑顔が出たときにかけられていた言葉のランキング等、受講生の心理状態により一層向き合うために必要な情報などが挙げられる。

[0083] 図示されるように、システムは、上述した解析部によって複数の観点（ニュートラル、幸福、驚き、不快、怒り、悲しみ、恐れ等）による評価結果をダッシュボードに一覧に表示する。各感情を象徴する表情アイコンと共に表示することで直感的に理解しやすいものとなる。表示は、例えば、一日のオンラインセッションの評価結果としてもよいし、週単位、月単位としてもよい。

[0084] ダッシュボードには、上記のほか、感情毎に最も強く表出した時の動画像のダイジェストや、その際の発言をテキスト化した情報を表示することとしてもよい。また、使用した言葉の頻度（口癖）を算出し、ランキングを表示することとしてもよい。

[0085] <第5の実施の形態>

図16を参照して、本発明の第5の実施の形態を説明する。本実施の形態によるシステムは、概略、当事者同士で行われる会話に不適切な表現があったかどうか（例えば、パワーハラスメントのような立場を利用した発言等）を発言された相手の反応を考慮して検知するものである。検知の方法としては、ルールベース（禁止キーワードの発言事実の特定と、相手側のネガティブ感情の表出）によるものと、機械学習的なアプローチなどが挙げられる。

[0086] 図示されるように、本システムは、上司と部下とのオンラインセッション等において、例えば上司ユーザの発した言葉にNGキーワード（上司の発言

として不適切な言葉）が組みられているかどうかを検出し、その際の部下の評価値を取得し、恐れや、不安、悲しみ、怒り、といったネガティブな感情の表出がその言葉を言われる前と比較して所定の範囲を超えて上昇していた場合には当該上司の発言を不適切な発言として記録する。不適切な発言は例えば、会社の人事部等に、そのダイジェスト動画像や、発言した言葉のテキスト等と共に通知されることとしてもよい。

[0087] ＜第6の実施の形態＞

図17及び図18を参照して、本発明の第6の実施の形態を説明する。本実施の形態によるシステムは、概略、動画像の発言を利用した所謂ワードクラウドである。システムは、取得した動画像の音声を認識し、認識した音声に含まれる単語をその発言頻度に応じたサイズに変換してテキスト表示する。

[0088] 表示する単語としては、評価対象動画像で使用された頻度の高い単語を抽出することとしてもよいし、評価対象動画像以外の動画像には含まれていなかった単語（今回の動画像特有の言葉）を抽出することとしてもよい。

[0089] また、システムは、当該単語を発した時のユーザの評価結果に応じてテキストの色を変更することとしてもよい。例えば、HAPPYスコアが高い場合には赤い文字とし、SADスコアが高いときに発言された単語は、青いスコアとしてもよい。

[0090] 図17に示されるワードクラウドは、中心に「勉強」という言葉が表示されている。当該単語を選択操作させると図18に示されるように、その単語を発言していたときの会話文がテキストで表示される。また、会話文と共に、再生ボタンPが表示され、再生ボタンPを選択すると当該テキストに対応する動画がダイジェストで再生される。

[0091] ＜第7の実施の形態＞

図19及び図20を参照して、本発明の第7の実施の形態を説明する。本実施の形態によるシステムは、概略、動画像に含まれる顔画像及び音声の双方に基づいて疲労度を評価するものである。

[0092] システムは、疲労度評価条件読込部と、疲労度判定部とを備えている。本実施の形態による疲労度の評価は、ユーザの定常状態を記憶し、当該定常状態における感情の起伏の幅と現在の感情の起伏の幅とに基づいて疲労度を評価する。なお、疲労度の評価は、これに限られず、定常状態における会話のピッチの変化量と、現在の会話のピッチの変化量とに基づいて評価してもよい。

[0093] 図20に示されるように、動画像を取得する（ステップS2000）と、予め学習させた疲労度評価モデルを読込み（ステップS2001）、疲労度の評価を行い（ステップS2002）、疲労度の通知（ステップS2010）を行う。又は、動画像を取得した後に通常時の評価情報を読込み（S2101）、各感情の評価結果と比較することにより各要素をクロス分析を行い疲労度を通知する（ステップS2010）。

[0094] このような疲労度の評価を行うことにより、例えば、従業員等に対して、疲労度に基づく段階的なアラートを通知することとしてもよい。例えば、疲労度が一定の閾値を超えた場合には、順次球威を促すメッセージを送信することとしてもよい。

[0095] <第8の実施の形態>

図21を参照して、本発明の第8の実施の形態を説明する。本実施の形態によるシステムは、概略、ユーザの発言の順序を考慮して時系列に可視化する。誰の後に誰が発言しやすいかといったことが分析可能となる。

[0096] システムは、動画像を取得し動画像内に含まれるユーザ毎に発言を認識する。システムは、記発言とユーザとを関連付けて時系列に並べて表示するオブジェクトを生成するオブジェクト生成部を備えている。

[0097] 図21は、ユーザA乃至Cの会話ラリーのオブジェクトを示す図である。発言がアッ場合には発言オブジェクトPがプロットされ、時間的に隣り合う発言オブジェクトP同士はコネクタCで接続される。

[0098] かかる会話ラリーから、例えば、ユーザCはユーザBの後に発言する傾向にあることが評価できる。この場合、例えば、ユーザCの発言を促すために

ユーザBの会話を増やすことなどが考えられる。

[0099] システムは、認識した顔画像及び発言の双方に基づいて所定の観点による評価値を算出する評価手段を更に備えている。会話ラリーの発言オブジェクトPには当該評価値に応じた色を付与することとしてもよい。例えば、ユーザが自然体で発言したあとに他のユーザが続いたのか、それとも、直前のユーザの高圧的な発言の後に別のユーザが続いたのか、では、改善方法も変わってくるからである。

[0100] <第9の実施の形態>

図22を参照して、本発明の第9の実施の形態を説明する。本実施の形態によるシステムは、概略、動画像内に含まれるユーザ毎に発言を認識し、当該発言に対応する発言オブジェクトをユーザと関連付けてプロットする発言オブジェクト生成部を備える。

[0101] 図示されるように、SUZUKI、SATO、KAMIYA、NOSE、ANDO

TADAの6名の発言時に発言オブジェクトがプロットされる。当該グラフを見ることにより、発言の有無を一見して理解することができる。

[0102] また、システムは、認識した顔画像及び発言の双方に基づいて所定の観点による評価値を算出する評価手段を更に備えている。発言オブジェクトPには当該評価値に応じた色を付与することとしてもよい。

[0103] かかる構成によれば、誰の発言量が多かったのか、全体としてどんな場だったのかが一見して容易に把握することができる。

[0104] <第10の実施の形態>

図23及び図24を参照して、本発明の第10の実施の形態を説明する。本実施の形態によるシステムは、概略、動画像から取得した音声の抑揚情報を抽出する抑揚取得手段と、認識した顔画像及び抑揚情報の双方に基づいて所定の観点による評価値を算出する評価手段とを備えている。抑揚取得手段は、単位時間当たりの音声の音程の変化を抽出することとしてもよい。

[0105] 図23は、音程（ピッチ）の標準偏差をとったグラフである。図24は、

音量の標準偏差をとったグラフである。本実施の形態においては、所定のサンプリングレートの音声データから、所定のフレーム数で標準偏差を取得している。

[0106] 一般に、コミュニケーションが良好な会話ほど、お互いの発話のピッチと音圧の標準偏差が大きいと言われている。このことから、標準偏差が全体的に低い値の会話は暗い会話である傾向があり、標準偏差の変化が少ない会話は淡々とした会話である傾向があり、時間を経過とともに標準偏差が単調減少すると時間と共に盛り下がった会話である傾向があり、時間を経過とともに標準偏差が減少するが最後増加に転じた場合は会話の最後は盛り上がった会話である傾向があり、標準偏差が高止まりしている場合は話者が焦っていたり混乱している傾向があると言える。このような音声の分析は、特に顔画像が映っていない動画像（カメラで取得した情報がない場合など）や、ユーザが下を向いたりして顔が写っていないシーンでも有効である。

[0107] 図23のt2及びt3は比較的単調であるが、t1及びt4は変化が大きいことがわかる。従って、t1及びt4のときには会話が盛り上がっていないことが推定できる。また、図24においても、図23と同じ時間軸に相当するt2及びt3は変化が少なく比較的単調であることがわかる。また、t1及びt4は比較的強弱が現れていることがわかる。このことから、音程及び音量の両方について、t2及びt3の時間帯は会話はあまり弾んでおらず、一方t1及びt4の時間帯は会話は弾んでいると推定できる。

[0108] <第11の実施の形態>

図25を参照して、本発明の第11の実施の形態を説明する。本実施の形態によるシステムは、概略、状況・情景などのコンテキストを理解して評価を行う。例えば、笑顔スコアが低くても、それが初対面だから低かったのか、親しい友人だけれど低かったのかの違いによって行うべき評価は異なるべきである。

[0109] 本システムは、コンテキスト情報を読み込むコンテキスト読込部と、当該コンテキスト情報に応じて評価結果を補正する補正部とを備えている。コン

テキスト情報は、例えば、シチュエーションや、会話を交わした回数、相手との面識、一方方向の会話形式又は双方向の会話形式等といったコンテキストを分類したカテゴリ情報と、補正すべき項目及び補正パラメータとを備えることとしてもよい。

[0110] システムは、ユーザからコンテキストのカテゴリ情報を事前に主導によって受け付けることとしてもよいし、動画像ファイル等のタイトルやメタデータから自動で判定することとしてもよい。これにより動画像のコンテキスト情報を特定し、該当するカテゴリに関連座けられている補正を行うことにより、適正な評価結果提供することができる。

[0111] 本明細書においてフローチャート図を用いて説明した処理は、必ずしも図示された順序で実行されなくてもよい。いくつかの処理ステップは、並列的に実行されてもよい。また、追加的な処理ステップが採用されてもよく、一部の処理ステップが省略されてもよい。

[0112] 以上説明した実施の形態を適宜組み合わせる実施することとしてもよい。また、本明細書に記載された効果は、あくまで説明的または例示的なものであって限定的ではない。つまり、本開示に係る技術は、上記の効果とともに、または上記の効果に代えて、本明細書の記載から当業者には明らかな他の効果を奏しうる。

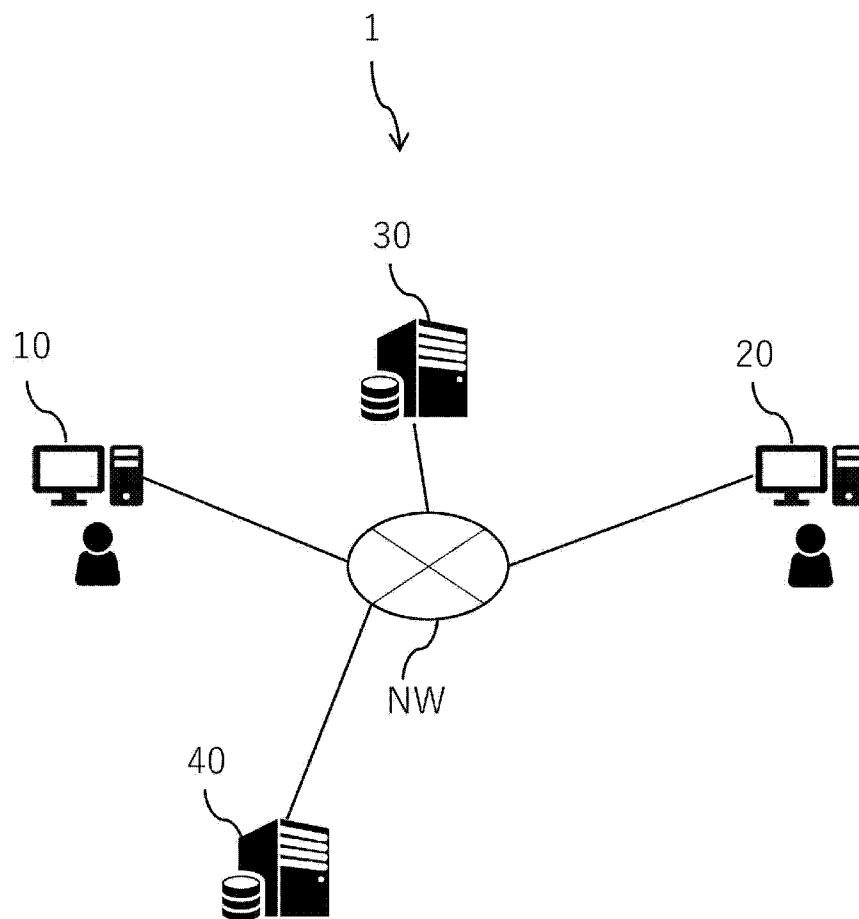
符号の説明

[0113] 10、20 ユーザ端末
 30 ビデオセッションサービス端末
 40 評価端末

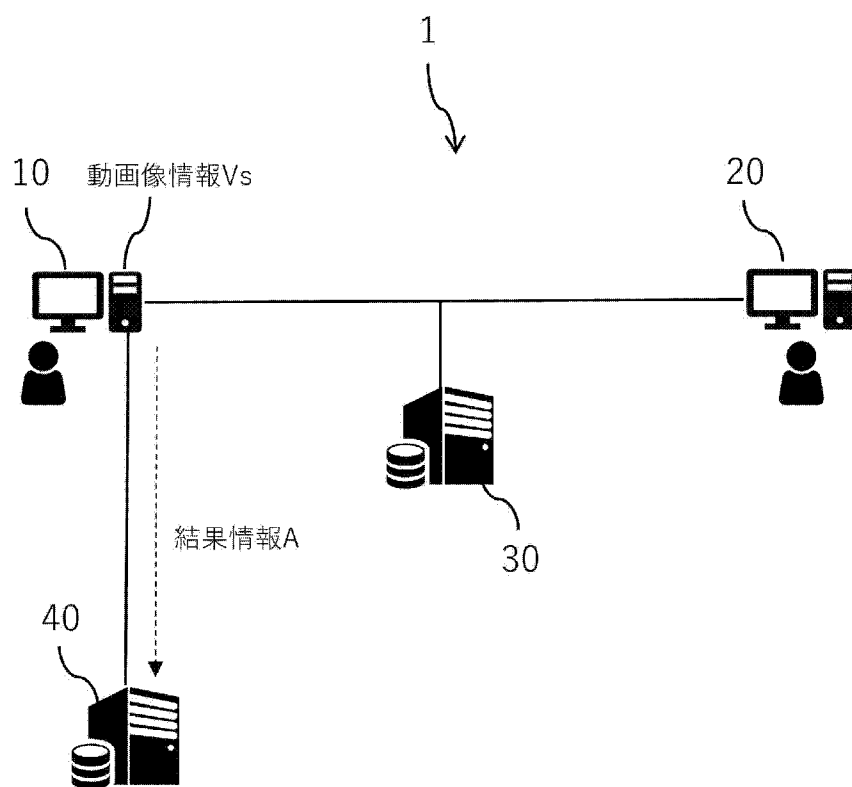
請求の範囲

- 【請求項1】 少なくとも動画像を取得する取得手段と、
- 前記動画像内に含まれる対象者の少なくとも顔画像を所定のフレームごとに認識する顔認識手段と、
- 前記動画像内に含まれる前記対象者の少なくとも音声を認識する音声認識手段と、
- 認識した前記音声の抑揚情報を抽出する抑揚取得手段と、
- 認識した前記顔画像及び前記抑揚情報の双方に基づいて所定の観点による評価値を算出する評価手段と、を備える
- 動画像分析システム。
- 【請求項2】 請求項1に記載の動画分析システムであって、
- 前記抑揚取得手段は、単位時間当たりの前記音声の音程の変化を抽出する、
- 動画像分析システム。
- 【請求項3】 請求項2に記載の動画分析システムであって、
- 前記抑揚取得手段は、単位時間当たりの前記音程の標準偏差を算出する、
- 動画像分析システム。

[図1]

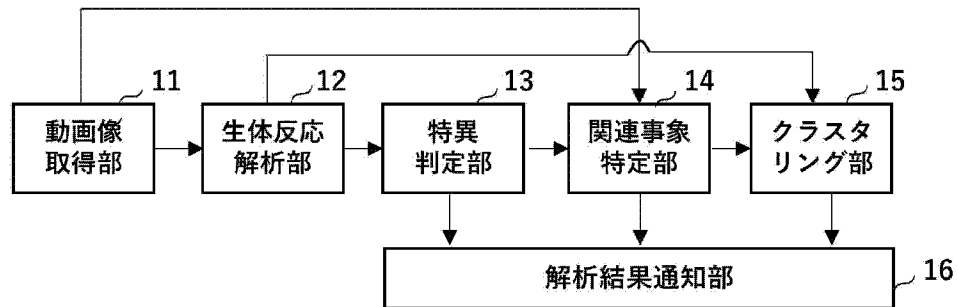


[図2]

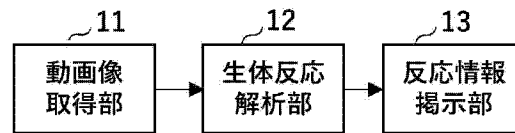


[図3]

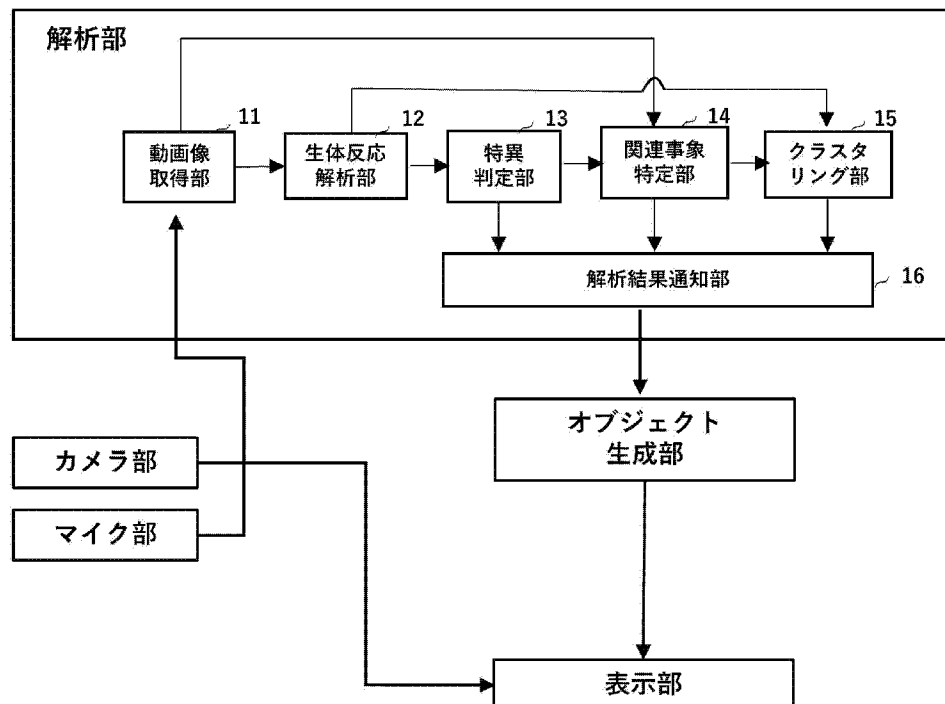
解析部の構成例



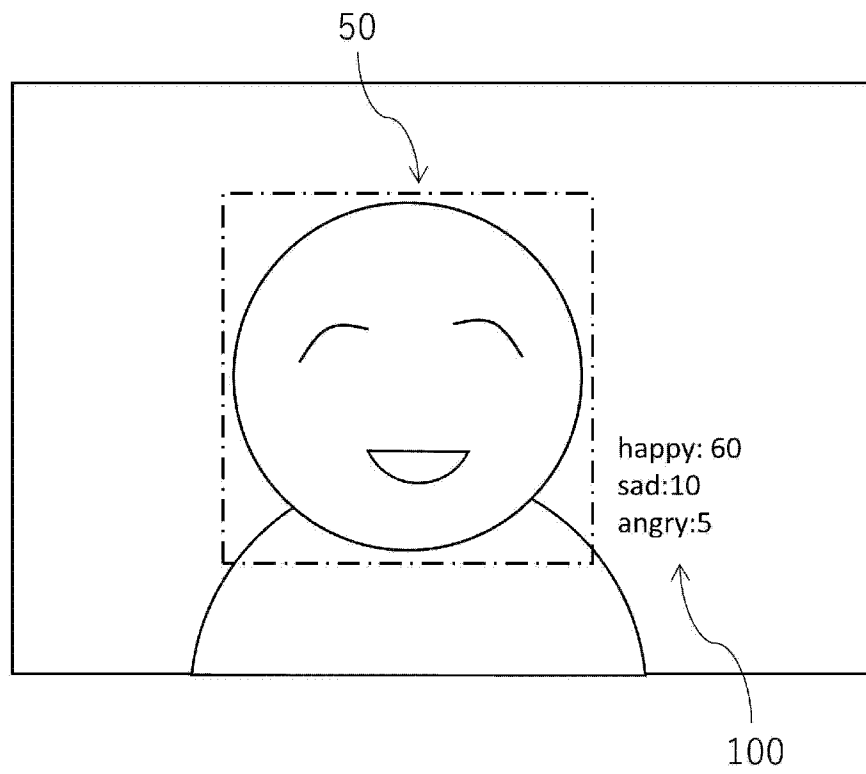
[図4]



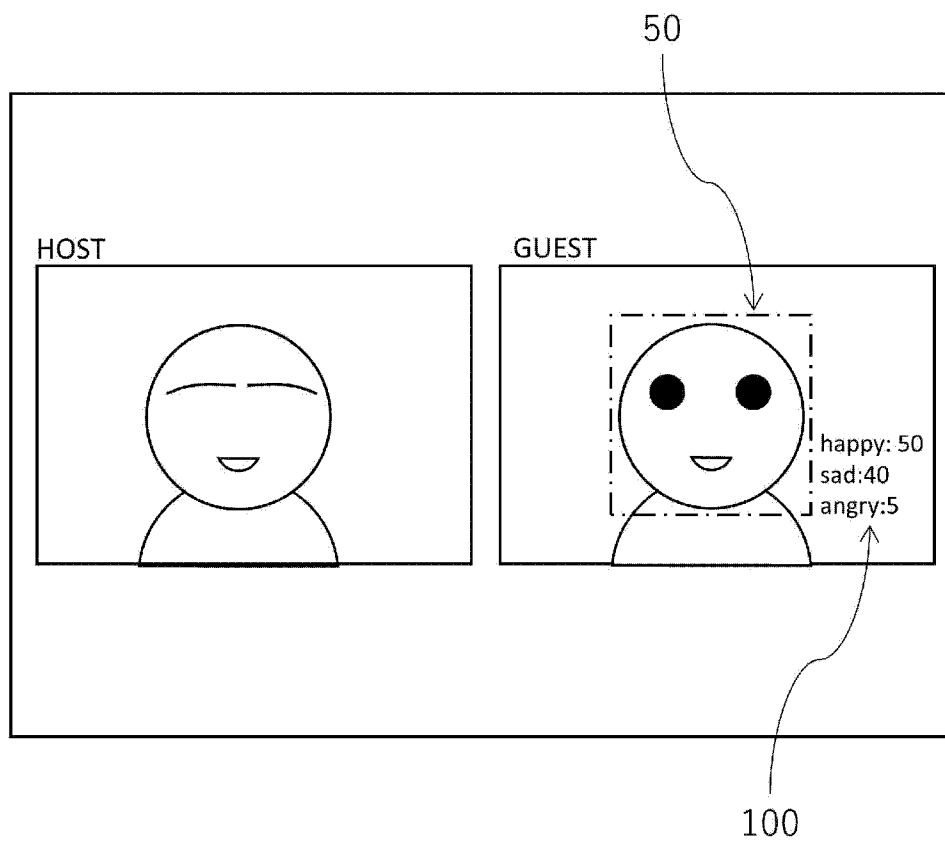
[図5]



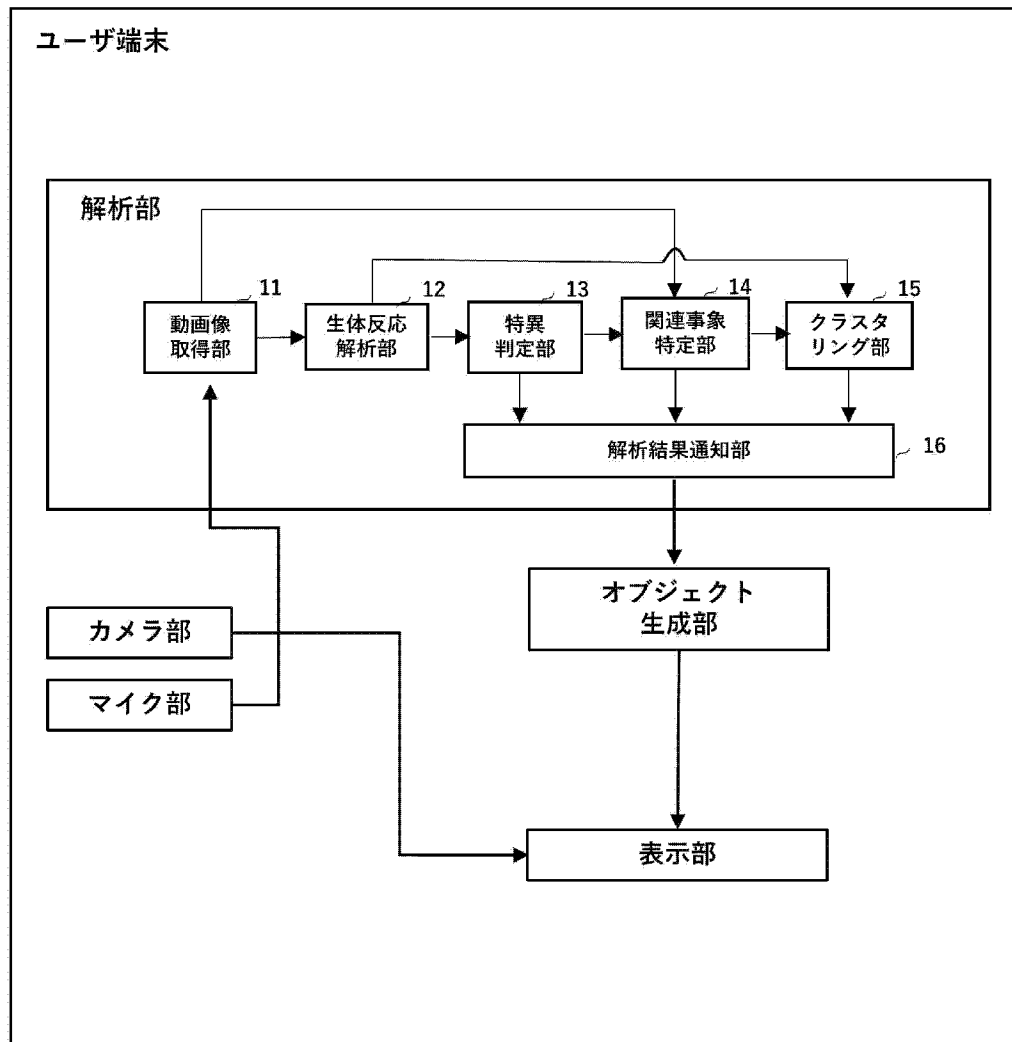
[図6]



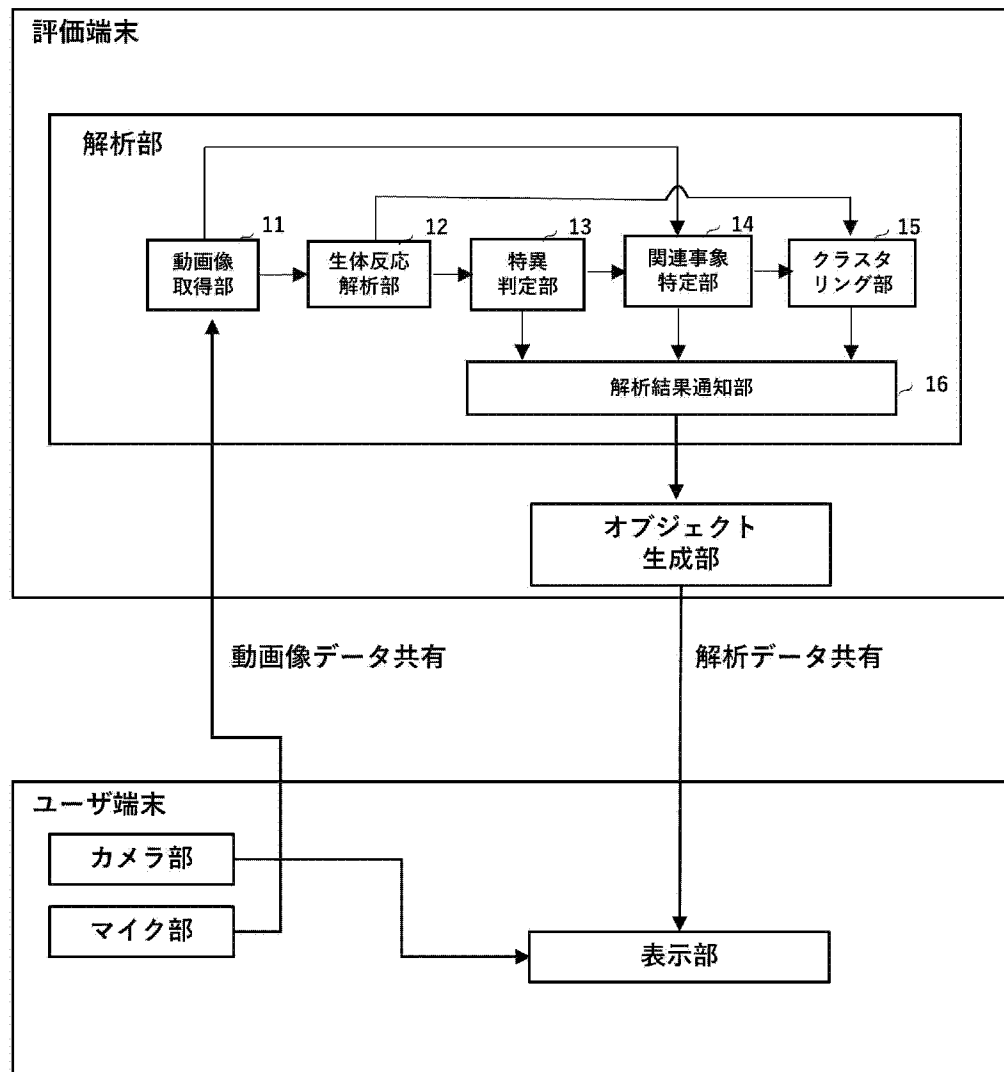
[図7]



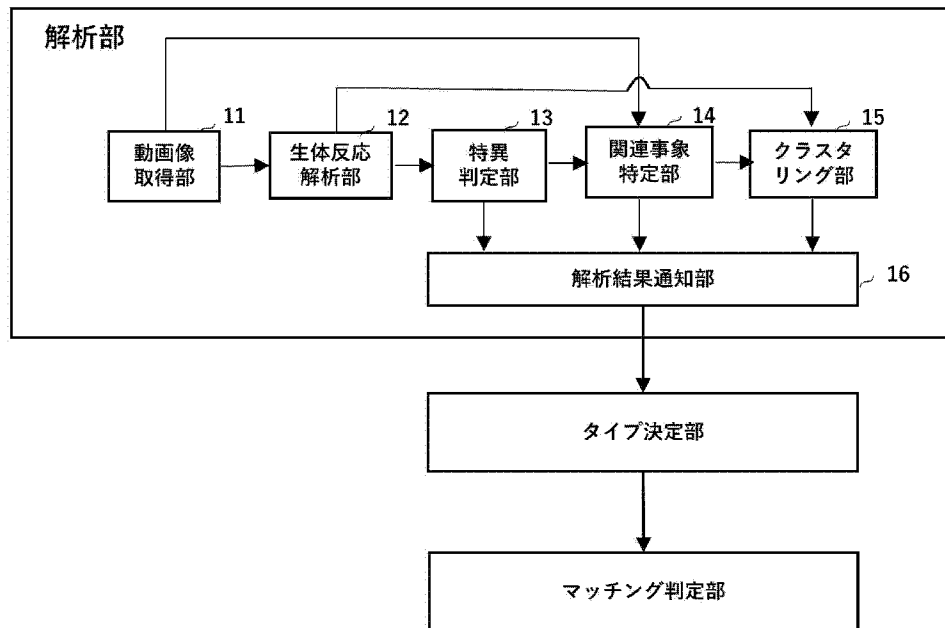
[図8]



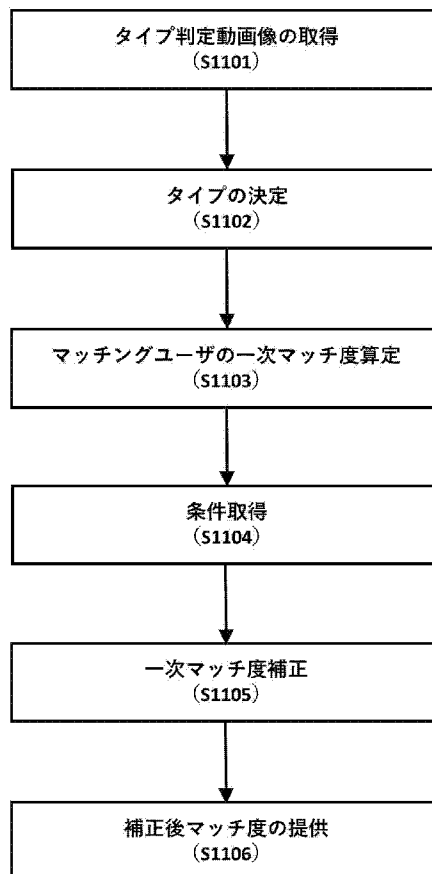
[図9]



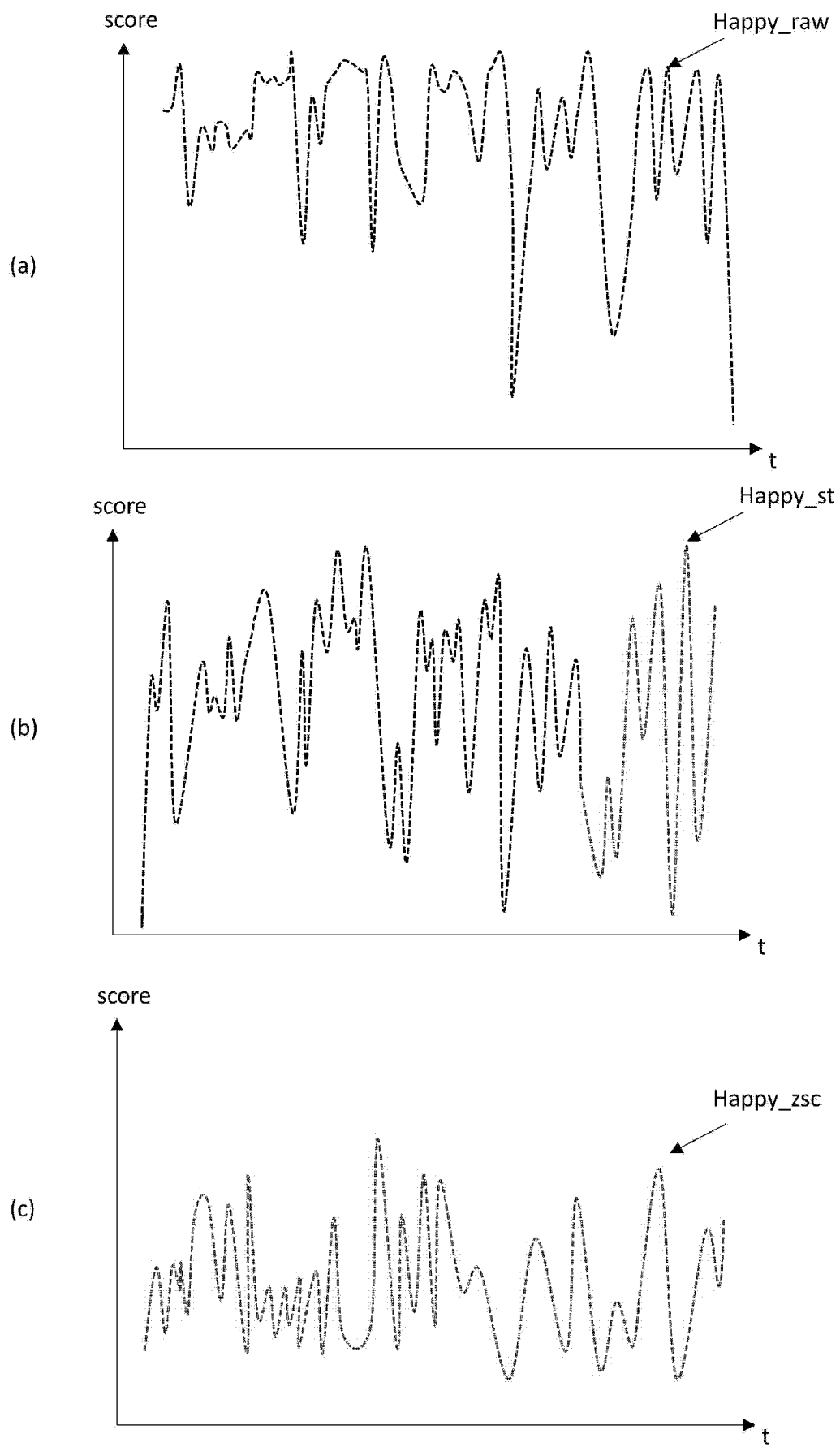
[図10]



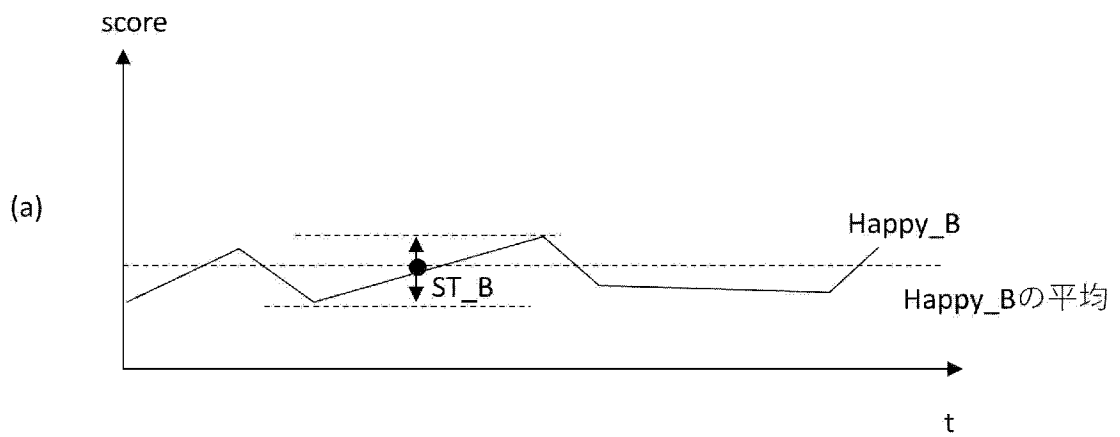
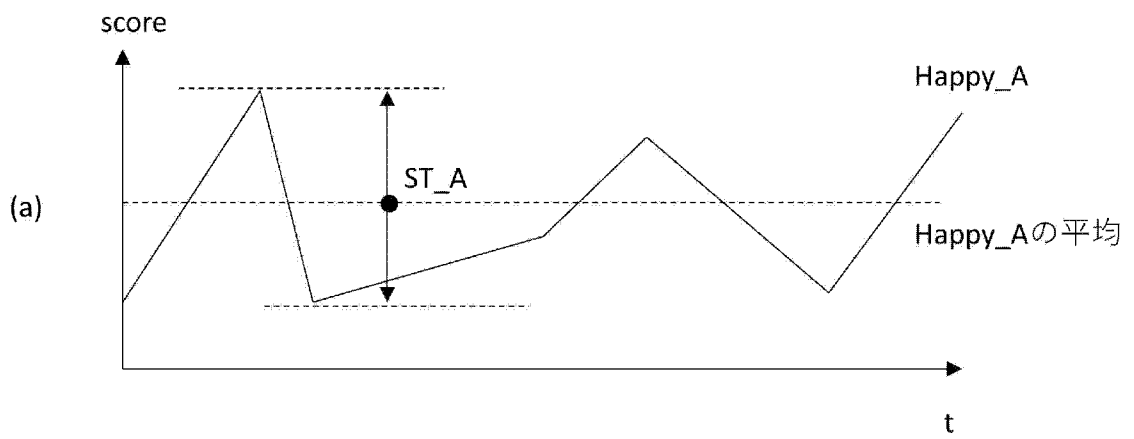
[図11]



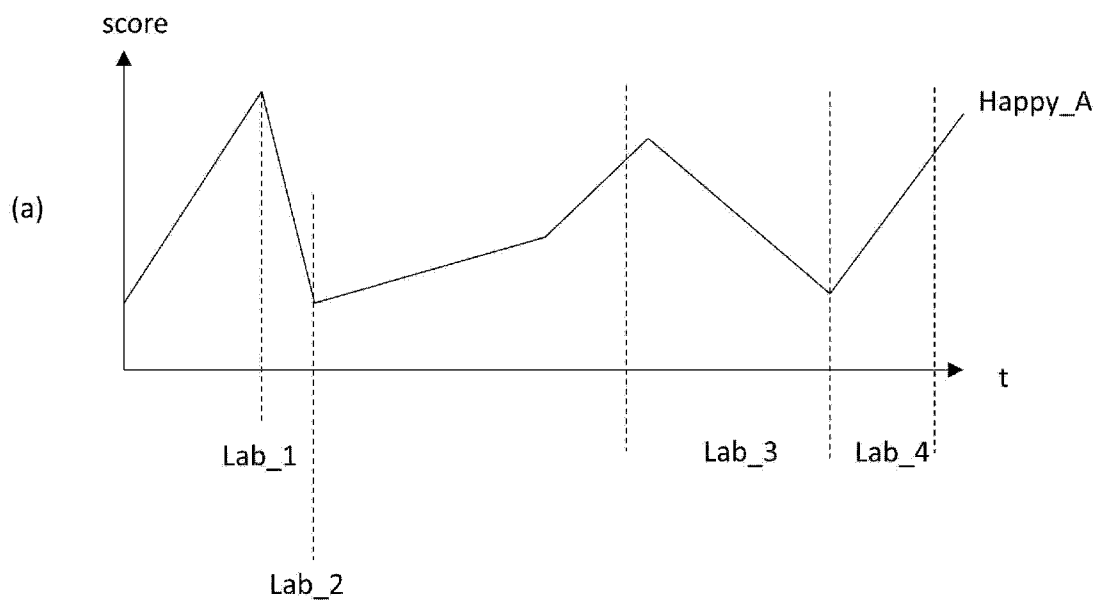
[図12]



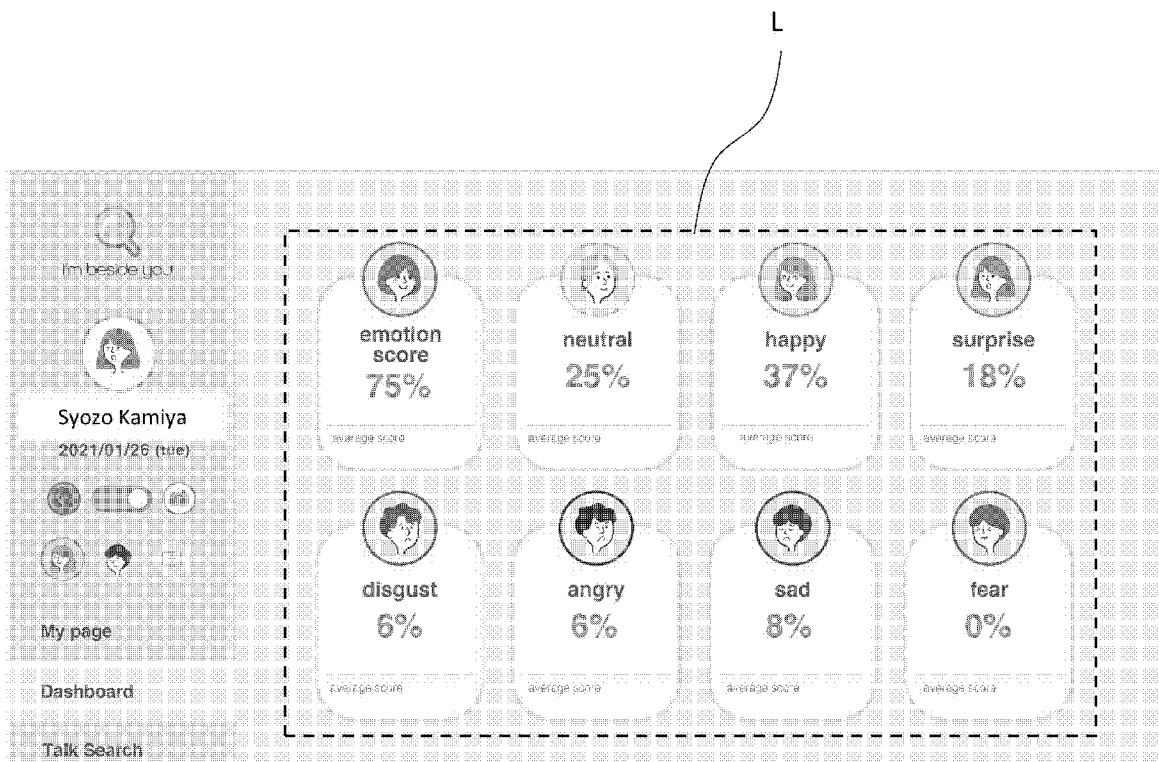
[図13]



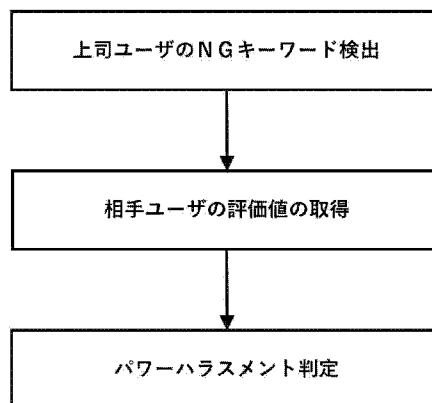
[図14]



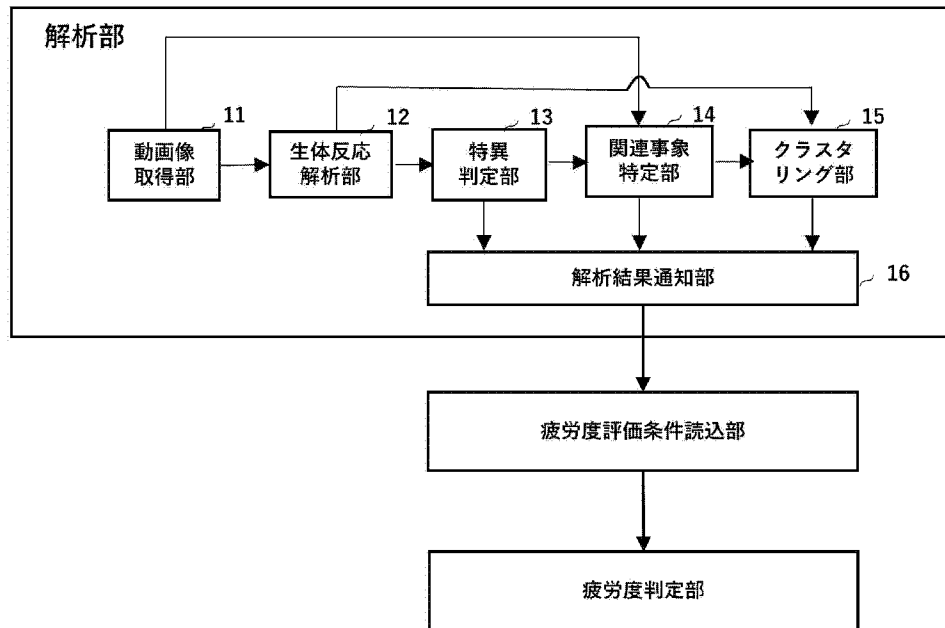
[図15]



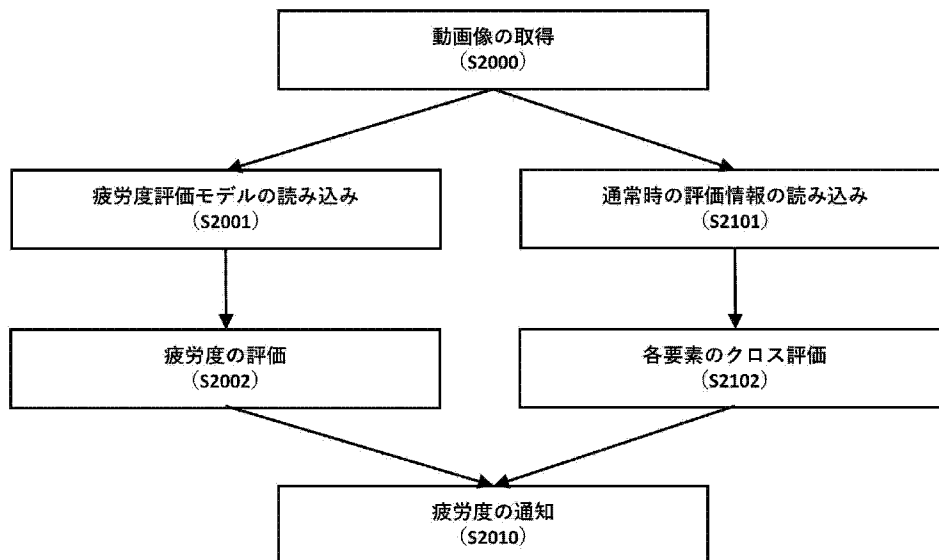
[図16]



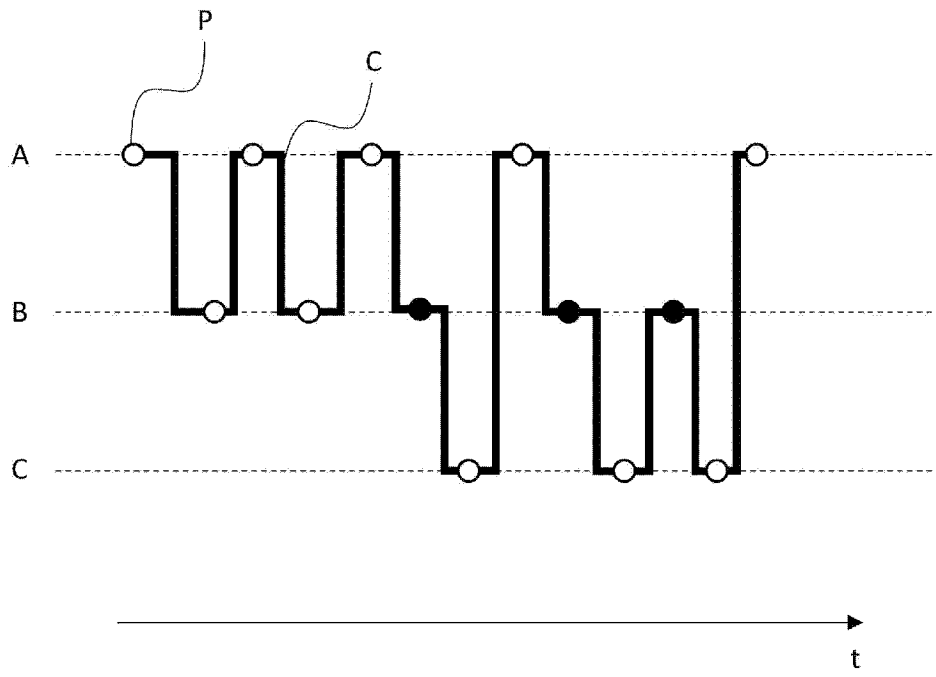
[図19]



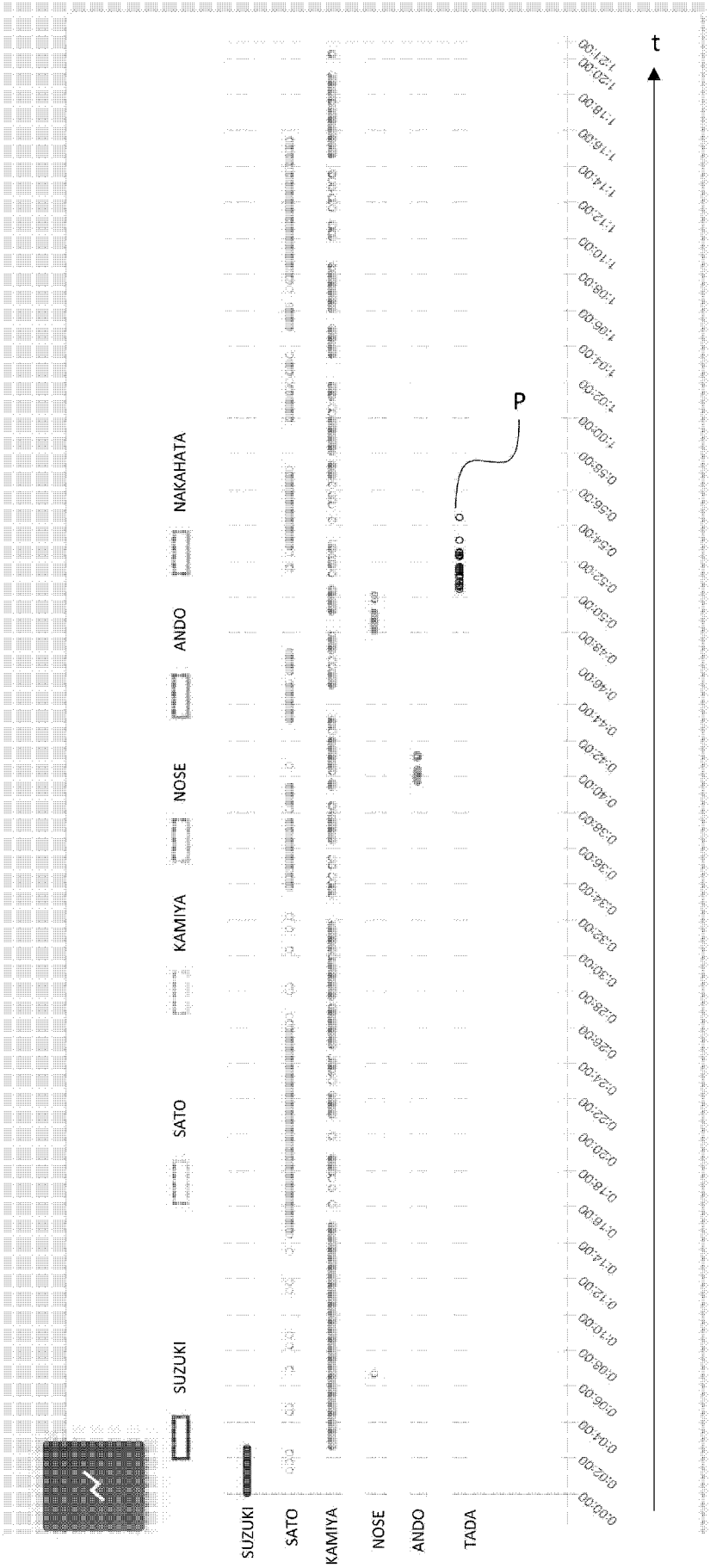
[図20]



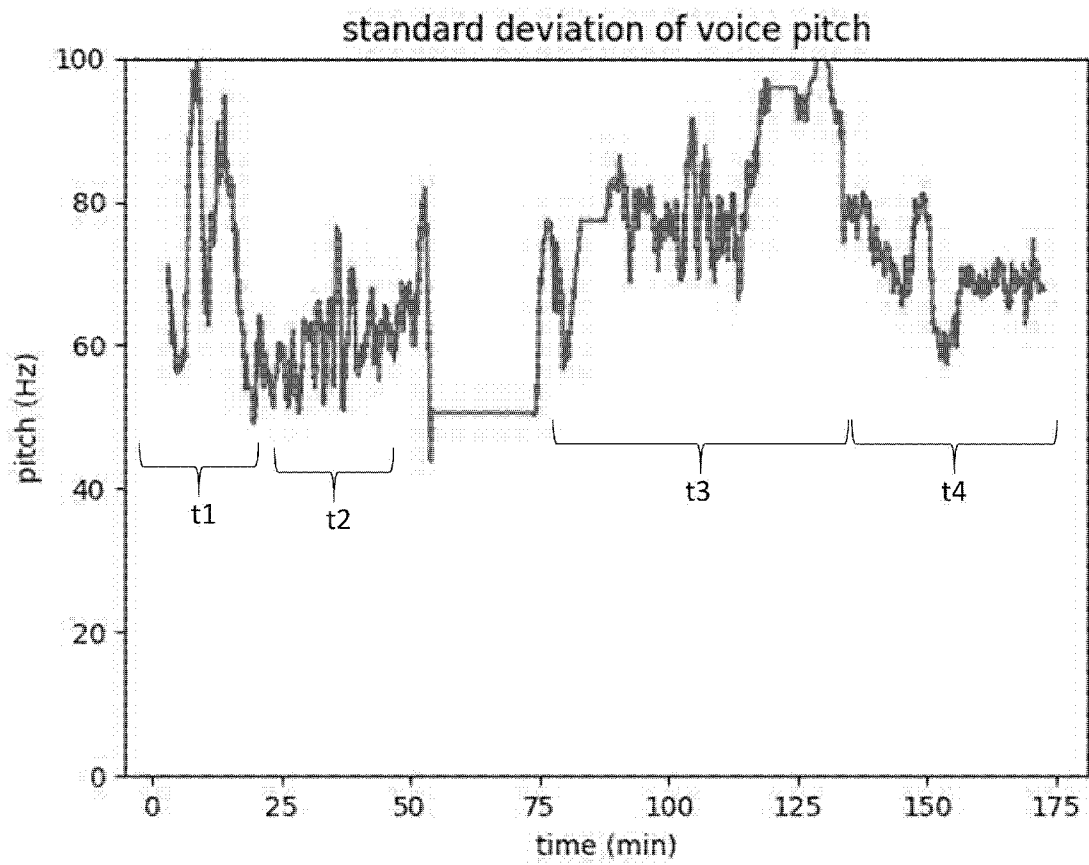
[図21]



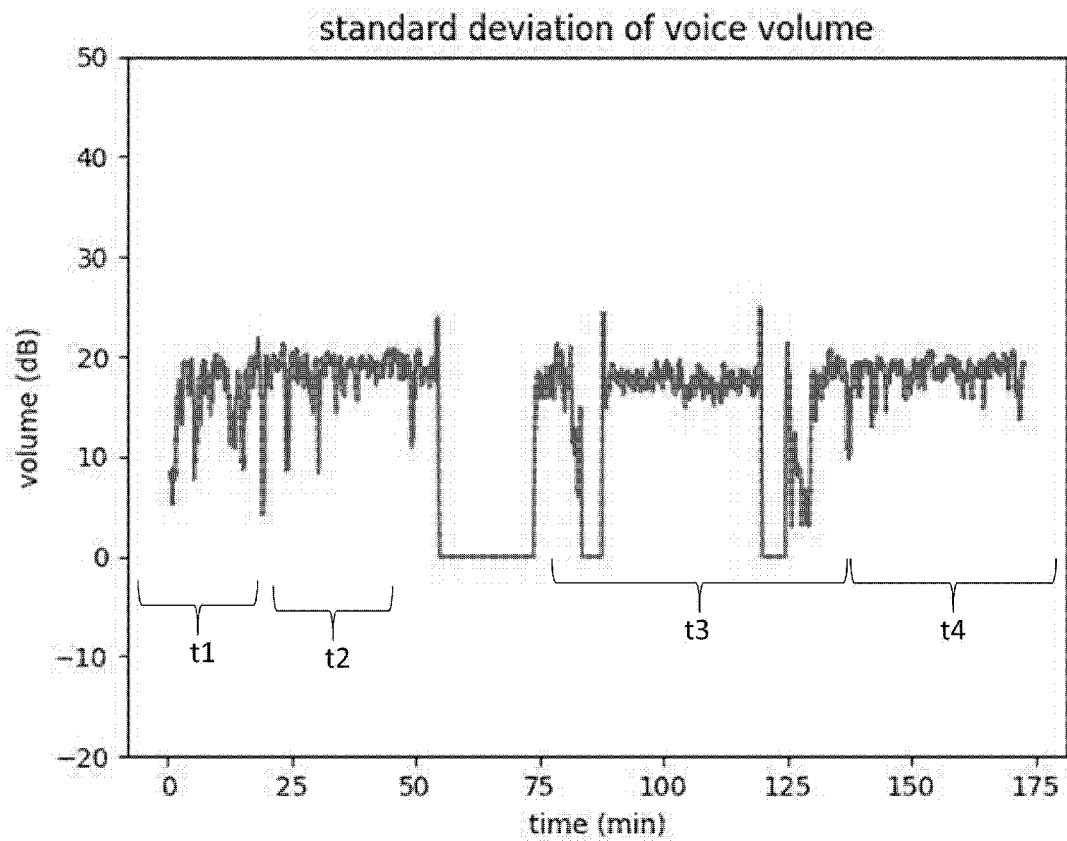
[図22]



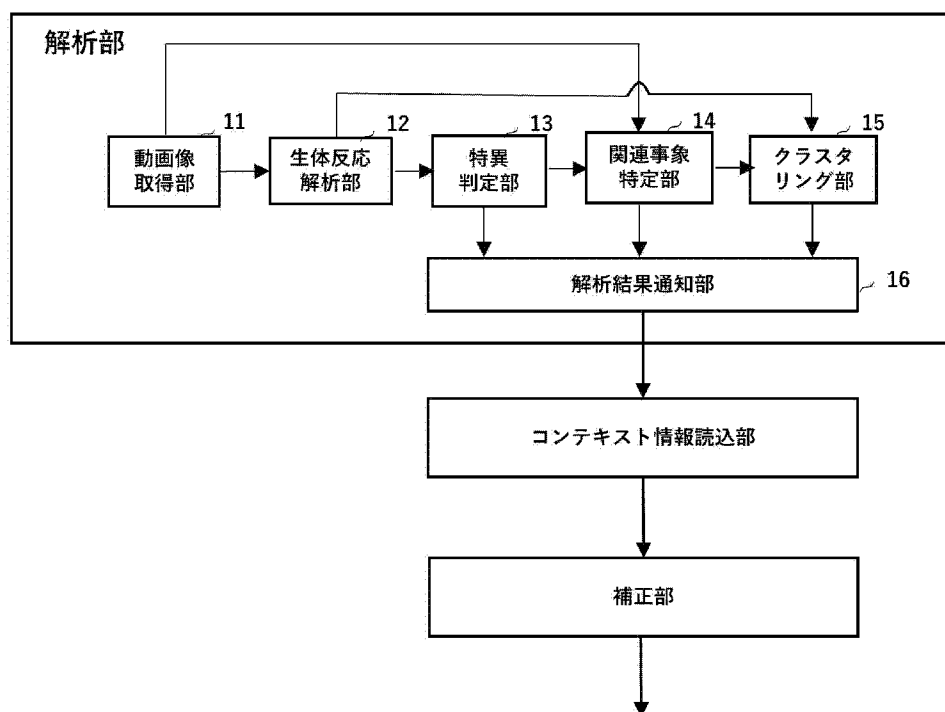
[図23]



[図24]



[図25]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2021/007578

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. H04N7/15 (2006.01) i

FI: H04N7/15

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. H04N7/15

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 1971-2021

Registered utility model specifications of Japan 1996-2021

Published registered utility model applications of Japan 1994-2021

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 Y	JP 2006-32062 A (SHARP KK) 30 November 2006 (2006-11-30), paragraphs [0011]-[0024], fig. 1, 5	1-2 3
X Y	JP 2005-348872 A (JAPAN BROADCASTING CORP.) 22 December 2005 (2005-12-22), paragraphs [0004], [0023], [0037], [0038], fig. 1	1-2 3
Y	JP 2003-508805 A (ACCENTURE LLP) 04 March 2003 (2003-03-04), paragraphs [0005], [0006]	3
Y	JP 2020-091302 A (HONDA MOTOR CO., LTD.) 11 June 2020 (2020-06-11), paragraphs [0071], [0072], fig. 9	3
Y	JP 2019-062490 A (OKI ELECTRIC IND CO., LTD.) 18 April 2019 (2019-04-18), paragraphs [0061], [0062], fig. 5	3



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

23 April 2021

Date of mailing of the international search report

11 May 2021

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
Information on patent family members

International application No.

PCT/JP2021/007578

JP 2006-32062 A	30 November 2006	(Family: none)
JP 2005-348872 A	22 December 2005	(Family: none)
JP 2003-508805 A	04 March 2003	US 6151571 A column 1, lines 39-56 WO 2001/016938 A1 EP 1770687 A1
JP 2020-091302 A	11 June 2020	US 2020/0175970 A1 paragraphs [0089], [0090], fig. 9 CN 111341349 A
JP 2019-062490 A	18 April 2019	(Family: none)

A. 発明の属する分野の分類（国際特許分類（IPC）） H04N 7/15(2006.01)i FI: H04N7/15		
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）） H04N7/15		
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 1971-2021年 日本国実用新案登録公報 1996-2021年 日本国登録実用新案公報 1994-2021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 2006-32062 A（シャープ株式会社）30.11.2006（2006-11-30） [0011]-[0024], 図1,5	1-2
Y		3
X	JP 2005-348872 A（日本放送協会）22.12.2005（2005-12-22） [0004], [0023], [0037]-[0038], 図1	1-2
Y		3
Y	JP 2003-508805 A（アクセンチュア リミテッド ライアビリティ パートナーシップ） 04.03.2003（2003-03-04） [0005]-[0006]	3
Y	JP 2020-091302 A（本田技研工業株式会社）11.06.2020（2020-06-11） [0071]-[0072], 図9	3
Y	JP 2019-062490 A（沖電気工業株式会社）18.04.2019（2019-04-18） [0061]-[0062], 図5	3
☐ 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、一般的な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抵触するものではなく、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 23.04.2021		国際調査報告の発送日 11.05.2021
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号		権限のある職員（特許庁審査官） 松元 伸次 5C 9563 電話番号 03-3581-1101 内線 3541

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2021/007578

引用文献			公表日	パテントファミリー文献	公表日
JP	2006-32062	A	30.11.2006	(ファミリーなし)	
JP	2005-348872	A	22.12.2005	(ファミリーなし)	
JP	2003-508805	A	04.03.2003	US 6151571 A	
				column1 line39-56	
				WO 2001/016938 A1	
				EP 1770687 A1	
JP	2020-091302	A	11.06.2020	US 2020/0175970 A1	
				[0089]-[0090], FIG.9	
				CN 111341349 A	
JP	2019-062490	A	18.04.2019	(ファミリーなし)	