(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
2019年9月19日(19.09.2019)

WIPO | PCT

(10) 国際公開番号
WO 2019/176590 A1

(51) 国際特許分類:
H04N 21/435 (2011.01)　*G10L 15/22* (2006.01)
G06F 3/0481 (2013.01)　*H04N 21/488* (2011.01)
G06F 3/16 (2006.01)　*H04N 21/84* (2011.01)
G10L 15/00 (2013.01)

(21) 国際出願番号：　PCT/JP2019/008140

(22) 国際出願日：　2019年3月1日(01.03.2019)

(25) 国際出願の言語：　日本語

(26) 国際公開の言語：　日本語

(30) 優先権データ：
特願 2018-048032　2018年3月15日(15.03.2018)　JP

(71) 出願人：ソニー株式会社 (**SONY CORPORATION**) [JP/JP]；〒1080075 東京都港区港南１丁目７番１号 Tokyo (JP).

(72) 発明者：山岸　靖明 (**YAMAGISHI, Yasuaki**)；〒1080075 東京都港区港南１丁目７番１号 ソニー株式会社内 Tokyo (JP).

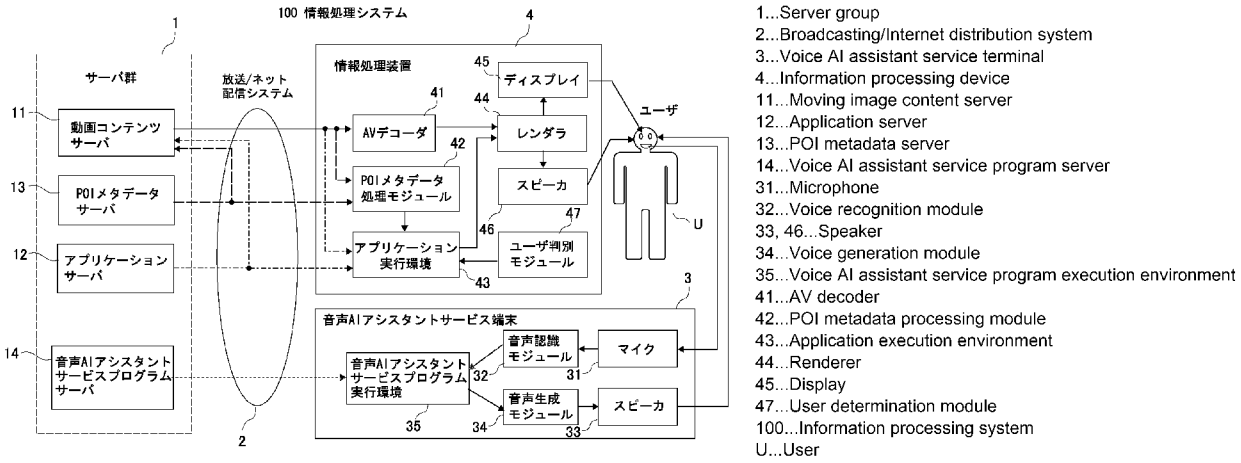
(74) 代理人：大森 純一(**OMORI, Junichi**)；〒1070052 東京都港区赤坂７－５－４７　Ｕ＆Ｍ赤坂ビル２Ｆ Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能)：AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

WO 2019/176590 A1

(54) **Title:** INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING DEVICE, AND PROGRAM

(54) 発明の名称： 情報処理装置、情報処理装置およびプログラム



(57) **Abstract:** This information processing device is provided with: an AV decoder 41 which acquires and reproduces video data including a service object capable of using a service for processing a voice request from a user; and an application execution environment 43 which adds, to reproduced video, an additional image for teaching the user the service object. The additional image has a visual characteristic unique to each service object such that the service object is uniquely determined by voice recognition in the service.

(57) 要約：この情報処理装置は、ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データを取得して再生するＡＶデコーダ４１と、再生した映像にサービス対象物をユーザに教示するための付加画像を付加するアプリケーション実行環境４３と、を具備する。付加画像は、サービス対象物がサービスにおいて音声認識によって一意に判別され得るように、サービス対象物毎にユニークな視覚的特徴を有する。

［続葉有］

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

添付公開書類:

一 国際調査報告(条約第21条(3))

明　細　書

発明の名称 ： 情報処理装置、情報処理装置およびプログラム

技術分野

[0001] 本技術は、映像を含む動画コンテンツを受信し、再生するための情報処理を行う情報処理装置、情報処理装置およびプログラムに関し、特にその動画コンテンツが情報処理装置のユーザに対して音声ベースによる情報サービスと連携するものである場合などに好適な情報処理装置、情報処理方法およびプログラムに関する。

背景技術

[0002] 近年、音声ＡＩアシスタントサービスが普及している。これは、情報処理装置のユーザが発した音声による要求を、サービスに対応する端末がマイクなどによって収音し、認識してデータを解析し、ユーザの要求に応じたサービスを実行して、その実行の結果をユーザに音声などで応答する情報サービスである（たとえば、特許文献１参照）。現在では、クラウドベースによる音声ＡＩアシスタンスサービスとして、Ａｍａｚｏｎ Ｅｃｈｏ（登録商標）のＡｌｅｘａ（登録商標）が知られている。

先行技術文献

特許文献

[0003] 特許文献1：特開２０１５－０２２３１０号公報

発明の概要

発明が解決しようとする課題

[0004] 上記のような音声ＡＩアシスタントサービスを、映像を含む動画コンテンツを再生する環境において、その映像に登場する人物や物品に関する情報を収集するための手段として利用する仕組みが本発明者らによって検討されている。例えば、動画コンテンツ中に登場する人物の役柄や、他の登場人物との関係、さらにはその人物を演じる俳優のプロフィールといった様々な事柄を視聴者であるユーザがその場で知りたい場合に、音声ＡＩアシスタントサ

ービスからリアルタイムに情報の提供を受けることができる。

[0005] しかしながら、音声ＡＩアシスタントサービスを、映像を含む動画コンテンツを再生する環境においてその映像に登場する人物や物品に関する情報を収集するための手段として実際に利用するとなると、音声ＡＩアシスタントサービスの特性を活かした好適な使い方が未だ十分確立されていないなど、解決すべき課題は山積されている。

[0006] そこで本技術は、再生映像を鑑賞する際に、音声ＡＩアシスタントサービスをその特性を活かして利用することのできる情報処理装置、情報処理装置およびプログラムを提供することを目的とするものである。

課題を解決するための手段

[0007] 上記の課題を解決するために、本技術に係る情報処理装置は、ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データを取得して再生するメディア再生部と、上記再生した映像に上記サービス対象物を上記ユーザに教示するための付加画像を付加する制御部とを具備する。

[0008] 上記付加画像は、上記サービス対象物が上記サービスにおいて音声認識によって一意に判別され得るように、前記サービス対象物毎にユニークな視覚的特徴を有するものであってよい。

上記付加画像は、上記サービス対象物に付随した位置に提示されてよい。

[0009] 上記制御部は、上記ユーザまたは上記ユーザの属性に応じて、前記付加画像が付加される前記サービス対象物を制限するように構成されたものであってよい。

[0010] 上記制御部は、上記ユーザまたは上記ユーザの属性に応じて上記付加画像が付加された上記サービス対象物が登場する期間の映像をスキップ再生するように構成されたものであってよい。

[0011] 上記制御部は、上記ユーザまたは上記ユーザの属性に応じて上記付加画像が付加された上記サービス対象物が登場する期間の映像を第１の倍速で再生し、上記サービス対象物が登場しない期間の映像を上記第１の倍速よりも高

速な第２の速度で再生するように構成されたものにしてもよい。

[0012] 上記制御部は、上記付加情報を生成するためのメタデータを取得し、上記取得したメタデータに基づき上記付加情報を付加するように構成されたものであってよい。

[0013] 上記制御部は、上記メタデータのＡｄａｐｔａｔｉｏｎＳｅｔを含むＭＰＤファイルを取得し、このＭＰＤファイルを解析して、上記映像データおよび上記メタデータをそれぞれＭＰＥＧ－ＤＡＳＨのＭｅｄｉａ　Ｓｅｇｍｅｎｔとして取得し、上記映像データと、上記メタデータに基づく上記付加画像とを互いに同期させて提示するように構成されたものであってよい。

[0014] 上記制御部は、上記メタデータのバージョン情報に基づいて、時間的に前後する前側のメタデータと後側のメタデータとの内容の変化の有無を判別し、変化がない場合、上記前側のメタデータに基づき映像に付加した付加画像を、上記後側のメタデータに基づく付加画像として、上記後側のメタデータに同期する映像に付加するように構成されてもよい。

[0015] また、上記付加画像の視覚的特徴は、上記サービス対象物に関する文字列、色、形状、またはアイコンのいずれか１つによって与えられるとしてよい。

[0016] 本技術に係る別の側面である情報処理方法は、ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データを取得して再生し、上記再生した映像に上記サービス対象物を上記ユーザに教示するための付加画像を付加するという手順を含むものである。

[0017] 本技術に係るさらに別の側面であるプログラムは、ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データをコンピュータにて再生するとき、前記再生した映像に前記サービス対象物を前記ユーザに教示するための付加画像を付加するようにコンピュータを機能させるプログラムである。

発明の効果

[0018] 以上のように、本技術によれば、再生映像を鑑賞する際に、音声ＡＩアシ

スタントサービスをその特性を活かして利用することができる。

図面の簡単な説明

[0019] [図1]図１は本技術に係る第１の実施形態の情報処理装置４を含む情報処理システム１００の全体構成を示すブロック図である。

[図2]図１の情報処理システム１００における全体的動作の流れ（その１）を示すシーケンス図である。

[図3]図１の情報処理システム１００における全体的動作の流れ（その２）を示すシーケンス図である。

[図4]図１の情報処理システム１００における全体的動作の流れ（その３）を示すシーケンス図である。

[図5]付加画像が重畳された映像の例を示す図である。

[図6]ＰＯＩメタデータの構成を示すブロック図である。

[図7]付加画像が重畳された映像の別の例を示す図である。

[図8]付加画像の提示制限を示す図である。

[図9]図１の情報処理システム１００における、付加画像の提示制限を含む全体的動作の流れ（その３）を示すシーケンス図である。

[図10]ＰＯＩメタデータに基づくトリックプレイ再生について説明する図である。

[図11]ＰＯＩメタデータを処理するアプリケーション実行環境４３の例を示す図である。

[図12]ＰＯＩメタデータを処理するアプリケーション実行環境４３の他の例を示す図である。

[図13]ウェブアプリケーションとＰＯＩメタデータをパッケージングするＭｕｌｔｉ－ｐａｒｔ　ＭＩＭＥフォーマットの例を示す図である

[図14]ＭＰ４ファイルのフォーマットにおけるＭｅｄｉａ　Ｓｅｇｍｅｎｔの構成を示す図である。

[図15]ＭＰＥＧ－ＤＡＳＨのＭＰＤのデータ構造を示す図である。

[図16]ＭＰＥＧ－ＤＡＳＨサーバ１５と情報処理装置４との間でのネットワ

ーク通信によるやりとりを示す図である。

[図17]ＭＰＥＧ－ＤＡＳＨ動画コンテンツの提示制御の流れを示す図である。

[図18]ＰＯＩメタデータのＡｄａｐｔａｔｉｏｎＳｅｔが追加されたＭＰＤの構成を示す図である。

[図19]ＰＯＩメタデータのＡｄａｐｔａｔｉｏｎＳｅｔが追加されたＭＰＤのより具体的な例を示す図である。

[図20]ＭＰＤに基づく映像および付加画像の提示の流れを示す図である。

[図21]サービス対象物の移動に伴って付加画像の提示位置を移動させる場合のＰＯＩメタデータを示す図である。

[図22]ＰＯＩメタデータの複数サンプル時間をまたがる提示更新制御を説明する図である。

発明を実施するための形態

[0020] 以下、本技術に係る実施形態を説明する。

[0021] ＜本実施形態の情報処理装置の要旨＞

本実施形態の情報処理装置４は、

ユーザからの音声による要求を処理する音声ＡＩアシスタントサービスを利用可能なサービス対象物を含む映像データを取得して再生するＡＶデコーダ４１と、

前記再生した映像に前記サービス対象物を前記ユーザに教示するための付加画像を付加するアプリケーション実行環境４３と、を具備する。

[0022] これにより、本実施形態の情報処理装置４は、再生映像において音声ＡＩアシスタントサービスのサービス対象物を一目で判別することができ、サービス対象物を選択する際に戸惑うことがなくなるという効果を有する。

[0023] ＜第１の実施形態＞

次に、第１の実施形態の情報処理装置４を含む情報処理システム１００の構成及びその動作を詳細に説明する。

[0024] （システム全体構成）

図１は本技術に係る第１の実施形態の情報処理装置４を含む情報処理システム１００の全体構成を示すブロック図である。

同図に示すように、この情報処理システム１００は、サーバ群１、放送／ネット配信システム２、音声ＡＩアシスタントサービス端末３、およびユーザの情報処理装置４を含む。

[0025] （サーバ群１について）

サーバ群１は、動画コンテンツサーバ１１、アプリケーションサーバ１２、ＰＯＩメタデータサーバ１３、および音声ＡＩアシスタントサービスプログラムサーバ１４を含む。

[0026] 動画コンテンツサーバ１１は、ＣＭや番組などの動画コンテンツのデータを蓄積する。動画コンテンツのデータは、映像、音声、字幕などの複数の種類のメディアデータで構成されるマルチメディアデータであってよい。本技術は、特に映像データに適用し得るものであり、音声、字幕などの他のメディアデータの有無は問わない。

[0027] ＰＯＩメタデータサーバ１３は、動画コンテンツに対応するＰＯＩメタデータを放送／ネット配信システム２を通じて情報処理装置４にダイレクトに配信する。あるいは、動画コンテンツサーバ１１において、少なくとも動画コンテンツのデータとＰＯＩメタデータを所定のデータ構造にまとめて情報処理装置４に配信する場合には、ＰＯＩメタデータサーバ１３はＰＯＩメタデータを放送／ネット配信システム２を通じて動画コンテンツサーバ１１に供給する。ＰＯＩメタデータについては後で説明する。

[0028] アプリケーションサーバ１２は、情報処理装置４においてＰＯＩメタデータに基づく処理を実行するためのアプリケーションを放送／ネット配信システム２を通じて情報処理装置４にダイレクトに配信する。あるいは、動画コンテンツサーバ１１において動画コンテンツのデータ、ＰＯＩメタデータおよびアプリケーションを所定のデータ構造にまとめて情報処理装置４に配信する場合には、アプリケーションサーバ１２はアプリケーションを放送／ネット配信システム２を通じて動画コンテンツサーバ１１に供給する。

[0029] 音声ＡＩアシスタントサービスプログラムサーバ１４は、放送／ネット配信システム２を通じて音声ＡＩアシスタントサービス端末３に音声ＡＩアシスタントサービスプログラムを供給する。音声ＡＩアシスタントサービスプログラムは、動画コンテンツに登場する特定の人物や物品などのサービス対象物について動画コンテンツの視聴者であるユーザＵから与えられたサービス要求に対するサービス実行結果を生成するように構成されたプログラムである。

[0030] 放送／ネット配信システム２は、動画コンテンツのデータ、あるいは動画コンテンツのデータとＰＯＩメタデータをまとめたデータ構造、あるいは動画コンテンツのデータとＰＯＩメタデータとアプリケーションをまとめたデータ構造を、放送あるいはネットワーク通信によって情報処理装置４に配信するシステムである。また、放送／ネット配信システム２は、動画コンテンツサーバ１１、アプリケーションサーバ１２、ＰＯＩメタデータサーバ１３の間でのデータ転送、さらには音声ＡＩアシスタントサービスプログラムサーバ１４と音声ＡＩアシスタントサービス端末３との間のデータ転送にも用いられる。

[0031] （情報処理装置４について）

情報処理装置４は、ハードウェア要素として、ＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）と、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）などのメインメモリと、ＨＤＤ（Ｈａｒｄ Ｄｉｓｋ Ｄｒｉｖｅ）、ＳＳＤ（Ｓｏｌｉｄ Ｓｔａｔｅ Ｄｒｉｖｅ）などのストレージデバイスと、ユーザインタフェースと、アンテナおよび放送チューナなどの放送受信部と、ネットワークインタフェースなどの通信インタフェースとを備える。情報処理装置４は、具体的には、パーソナルコンピュータ、スマートホン、タブレット端末、テレビジョン、ゲーム機、ＨＭＤ（Ｈｅａｄ Ｍｏｕｎｔｅｄ Ｄｉｓｐｌａｙ）などのユーザ装着可能形の情報端末などであってよい。

[0032] 情報処理装置４は、上記のハードウェア要素とソフトウェア要素によって

実現される機能的要素として、ＡＶデコーダ４１、ＰＯＩメタデータ処理モジュール４２、アプリケーション実行環境４３、レンダラ４４を有する。

[0033] 情報処理装置４は、ユーザインタフェースとして、ディスプレイ４５とスピーカ４６を備える。さらに、情報処理装置４は、機能的な要素としてユーザ判別モジュール４７を備えたものであってよい。

[0034] ＡＶデコーダ４１は、放送／ネット配信システム２を通じて動画コンテンツサーバ１１より取得した動画コンテンツのマルチメディアデータ（映像データおよび音声データなど）を復号する。

[0035] ＰＯＩメタデータ処理モジュール４２は、放送／ネット配信システム２を通じて動画コンテンツサーバ１１あるいはＰＯＩメタデータサーバ１３より取得したＰＯＩメタデータをアプリケーション実行環境４３に供給する。

[0036] 「ＰＯＩメタデータ」とは、情報処理装置４において再生中の映像に、当該映像中で音声ＡＩアシスタントサービスが割り当てられた特定のサービス対象物を、情報処理装置４のユーザＵに教示する追加画像を追加するためのメタデータである。ＰＯＩはＰｏｉｎｔ　ｏｆ　Ｉｎｔｅｒｅｓｔの略である。「特定のサービス対象物」は、映像において音声ＡＩアシスタントサービスがユーザＵからのサービス要求に対して応答することが可能とされた登場人物や物品などのことである。

[0037] アプリケーション実行環境４３は、ネイティブアプリケーションまたはウェブアプリケーションをＣＰＵおよびメインメモリを用いて実行する環境である。アプリケーション実行環境４３は、ＰＯＩメタデータ処理モジュール４２より与えられたＰＯＩメタデータに基づいて、映像中のサービス対象物に付加すべき追加画像を生成する。「追加画像」は例えば、映像中のサービス対象物との関係が分かりやすいように、例えば「吹き出し」として生成される。しかし、本技術はこれに限定されず、映像中のサービス対象物との関係が分かりやすい画像であればよい。

[0038] レンダラ４４は、ＡＶデコーダ４１により復号された映像データからディスプレイ４５に出力する表示信号を生成したり、ＡＶデコーダ４１により復

号された音声データをスピーカ４６に出力する。また、レンダラ４４は、アプリケーション実行環境４３より追加画像が供給された場合にはこの追加画像を番組映像の上に合成する。

[0039] ディスプレイ４５はユーザＵに映像を提示する。スピーカ４６はユーザＵに音声を提示する。

[0040] （音声ＡＩアシスタントサービス端末３について）

音声ＡＩアシスタントサービス端末３は、情報処理装置４のユーザＵに対し、音声によるアシスタントサービスを提供する端末である。音声ＡＩアシスタントサービス端末３は、より具体的には、映像中でユーザＵより任意のサービス対象物についての音声によるサービス要求を受け付け、そのサービスを実行し、サービスの実行結果をユーザＵに音声などで返すことのできる装置である。ここで、ユーザＵからの音声によるサービス要求は例えば質問形式などの言葉により与えられ、サービスの実行結果は例えば回答形式などの合成音声によってユーザＵに返される。

[0041] 音声ＡＩアシスタントサービス端末３は、アシスタントサービス用のマイク３１と、音声認識モジュール３２と、音声ＡＩアシスタントサービス用のスピーカ３３と、音声生成モジュール３４と、音声ＡＩアシスタントサービスプログラム実行環境３５を備える。

[0042] アシスタントサービス用のマイク３１は、情報処理装置４のユーザＵからの音声によるサービス要求を取り込む。

[0043] 音声認識モジュール３２は、アシスタントサービス用のマイク３１により取り込んだサービス要求の音声を認識して要求データを音声ＡＩアシスタントサービスプログラム実行環境３５にわたす。

[0044] 音声ＡＩアシスタントサービスプログラム実行環境３５は、音声ＡＩアシスタントサービスプログラムサーバ１４より取得した音声ＡＩアシスタントサービスプログラムを実行する環境である。音声ＡＩアシスタントサービスプログラム実行環境３５は、音声認識モジュール３２より供給された要求データに対するサービスの実行結果のデータを生成し、音声生成モジュール３

４に供給する。

[0045] 音声生成モジュール３４は、音声ＡＩアシスタントサービスプログラム実行環境３５より供給されたサービス実行結果のデータを合成音声に変換する。

[0046] アシスタントサービス用のスピーカ３３は、音声生成モジュール３４より供給された合成音声を情報処理装置４のユーザＵに提示する。

[0047] （情報処理システムの全体的動作の流れ（その１））

図２は、本実施形態の情報処理システム１００における全体的動作の流れ（その１）を示すシーケンス図である。

前提として、動画コンテンツのデータ、アプリケーションおよびＰＯＩメタデータがそれぞれ別々のサーバ（動画コンテンツサーバ１１、アプリケーションサーバ１２、ＰＯＩメタデータサーバ１３）から配信される場合を想定している。

[0048] まず、アプリケーションサーバ１２から情報処理装置４に放送／ネット配信システム２を通じてアプリケーションが配信される（ステップＳ１０１）。さらに動画コンテンツサーバ１１から情報処理装置４に動画コンテンツのデータが放送／ネット配信システム２を通じて配信される（ステップＳ１０２）。情報処理装置４において、受信した動画コンテンツのデータはＡＶデコーダ４１にて復号され、この結果得られた映像データおよび音声データがレンダラ４４を通じてディスプレイ４５及びスピーカ４６にそれぞれ供給される（ステップＳ１０３）。

[0049] また、アプリケーションサーバ１２から情報処理装置４に配信されたアプリケーションはアプリケーション実行環境４３に導入され、アプリケーションが実行される（ステップＳ１０４）。

[0050] 次に、ＰＯＩメタデータサーバ１３から情報処理装置４に放送／ネット配信システム２を通じて動画コンテンツに対応するＰＯＩメタデータが配信される（ステップＳ１０５）。情報処理装置４において、ＰＯＩメタデータはＰＯＩメタデータ処理モジュール４２によってアプリケーション実行環境４

３に供給される。アプリケーション実行環境４３は、ＰＯＩメタデータに基づいて、映像中のサービス対象物をユーザＵに教示する付加画像を生成し、レンダラ４４に供給する。これにより、番組の映像の上に付加画像が重畳された合成像が得られ、ディスプレイ４５に表示される（ステップＳ１０６）。

[0051] その後、ユーザＵは、ディスプレイ４５に表示された映像中で付加画像が提示されたサービス対象物についての音声によるサービス要求を音声ＡＩアシスタントサービス端末３に与えて、音声ＡＩアシスタントサービスを利用する（ステップＳ１０７）。

[0052] （情報処理システムの全体的動作の流れ（その２））

図３は本実施形態の情報処理システム１００における全体的動作の流れ（その２）を示すシーケンス図である。

前述の動作の流れ（その１）では、動画コンテンツのデータとＰＯＩメタデータがそれぞれ別々のサーバ（動画コンテンツサーバ１１、ＰＯＩメタデータサーバ１３）から情報処理装置４に配信されるため、情報処理装置４において動画コンテンツデータとＰＯＩメタデータとの同期が保証されない。

[0053] そこで、図３に示す動作の流れ（その２）では、まず、ＰＯＩメタデータサーバ１３から動画コンテンツサーバ１１に放送／ネット配信システム２を通じてＰＯＩメタデータが供給される（ステップＳ２０１）。続いて、動画コンテンツサーバ１１にて、ＰＯＩメタデータと動画コンテンツのデータとが所定のデータ構造にまとめられて情報処理装置４に放送／ネット配信システム２を通じて配信される（ステップＳ２０３）。

[0054] 情報処理装置４では、受信したデータ構造から動画コンテンツのデータが抽出され、抽出された動画コンテンツのデータはＡＶデコーダ４１にて復号され、この結果得られた映像データおよび音声データがレンダラ４４を通じてディスプレイ４５およびスピーカ４６にそれぞれ供給される（ステップＳ２０４）。

[0055] また、情報処理装置４では、受信したデータ構造からＰＯＩメタデータ処

理モジュール４２によってＰＯＩメタデータが抽出され（ステップＳ２０５）、アプリケーション実行環境４３に供給される。アプリケーション実行環境４３は、ＰＯＩメタデータに基づいて、映像中のサービス対象物をユーザＵに教示する付加画像を生成し、レンダラ４４に供給する。これにより、番組の映像の上に付加画像が重畳された合成像が得られ、ディスプレイ４５に表示される（ステップＳ２０７）。

[0056] その後、ユーザＵは、ディスプレイ４５に表示された映像中で付加画像が提示されたサービス対象物についての音声によるサービス要求を音声ＡＩアシスタントサービス端末３に与えて、音声ＡＩアシスタントサービスを利用する（ステップＳ２０８）。

[0057] このように、ＰＯＩメタデータと動画コンテンツのデータとを所定のデータ構造にまとめて動画コンテンツサーバ１１から情報処理装置４に配信することによって、情報処理装置４において動画コンテンツのデータとＰＯＩメタデータとを互いに同期して処理できることが保証される。このため、番組の映像中のサービス対象物に常に正しい付加画像を付加することができ、安定した音声ＡＩアシスタントサービスを維持することができる。

[0058] なお、この動作の流れ（その２）において、アプリケーションは、アプリケーションサーバ１２から情報処理装置４に放送／ネット配信システム２を通じて配信され（ステップＳ２０２）、アプリケーション実行環境４３に導入されてアプリケーションが実行される（ステップＳ２０６）という点は、前述の動作の流れ（その１）と同じである。

[0059] （情報処理システムの全体的動作の流れ（その３））

図４は本実施形態の情報処理システム１００における全体的動作の流れ（その３）を示すシーケンス図である。

この動作の流れでは、動画コンテンツサーバ１１から情報処理装置４に、動画コンテンツのデータとＰＯＩメタデータとこれを処理するアプリケーションが所定のデータ構造にまとめて配信される場合を想定している。

[0060] まず、ＰＯＩメタデータサーバ１３から動画コンテンツサーバ１１にＰＯ

Ｉメタデータが供給される（ステップＳ３０１）。さらに、アプリケーションサーバ１２から動画コンテンツサーバ１１にアプリケーションが供給される（ステップＳ３０２）。ＰＯＩメタデータの供給とアプリケーションの供給の順位は逆であってもよい。続いて、動画コンテンツサーバ１１にて、動画コンテンツのデータとＰＯＩメタデータとアプリケーションとが所定のデータ構造にまとめられて情報処理装置４に放送／ネット配信システム２を通じて配信される（ステップＳ３０３）。

[0061] 情報処理装置４では、受信したデータ構造から動画コンテンツのデータが抽出され、抽出された動画コンテンツのデータはＡＶデコーダ４１にて復号され、この結果得られた映像データおよび音声データがレンダラ４４を通じてディスプレイ４５およびスピーカ４６にそれぞれ供給される（ステップＳ３０４）。

[0062] また、情報処理装置４では、受信したデータ構造からアプリケーションが抽出され（ステップＳ３０５）、アプリケーション実行環境４３に導入されてアプリケーションが実行される（ステップＳ３０６）。

[0063] さらに、情報処理装置４では、受信したデータ構造からＰＯＩメタデータ処理モジュール４２によってＰＯＩメタデータが抽出され（ステップＳ３０７）、アプリケーション実行環境４３に供給される。アプリケーション実行環境４３は、ＰＯＩメタデータに基づいて、映像中のサービス対象物をユーザＵに教示する付加画像を生成し、レンダラ４４に供給する。これにより、番組の映像の上に付加画像が重畳された合成像が得られ、ディスプレイ４５に表示される（ステップＳ３０８）。

[0064] その後、ユーザＵは、ディスプレイ４５に表示された映像中で付加画像が提示されたサービス対象物についての音声によるサービス要求を音声ＡＩアシスタントサービス端末３に与えて、音声ＡＩアシスタントサービスを利用する（ステップＳ３０９）。

[0065] （付加画像について）

次に、ＰＯＩメタデータに基づき生成される付加画像について説明する。

図５は付加画像が重畳された映像の例を示す図である。

同図に示すように、付加画像は、例えば、映像中の音声ＡＩアシスタントサービスのサービス対象物Ｊ１、Ｊ２、Ｊ３にそれぞれ付随した吹き出し５１、５２、５３として提示されてよい。

[0066] 吹き出し５１、５２、５３には、サービス対象物Ｊ１、Ｊ２、Ｊ３をユーザＵが音声で音声ＡＩアシスタントサービス端末３に通知する場合にこのサービス対象物Ｊ１、Ｊ２、Ｊ３が音声ＡＩアシスタントサービス端末３において音声認識によって一意に判別され得るように、サービス対象物Ｊ１、Ｊ２、Ｊ３毎にユニークな視覚的特徴が与えられる。これによりユーザＵは、任意のサービス対象物の特徴を表現する音声を使って、そのサービス対象物についてのサービス要求を音声ＡＩアシスタントサービス端末３に与えることができる。

[0067] 例えば、図５の例では、吹き出し５１、５２、５３の中に”Ｂｉｌｌ”、”Ｆｒｅｄ”、”Ｌｕｃｙ”といったサービス対象物Ｊ１、Ｊ２、Ｊ３である登場人物の名前の文字列が視覚的特徴として表示される。これによりユーザＵは、例えば、”Ｆｒｅｄの最近のニュースは？”、”Ｂｉｌｌは何歳？”のように、任意のサービス対象物の名前を使って当該サービス対象物についてのサービス要求を音声ＡＩアシスタントサービス端末３に音声で与えることができる。一方、音声ＡＩアシスタントサービス端末３は音声認識モジュール３２によって認識された名前からサービス対象物を特定することができ、特定したサービス対象物について、音声ＡＩアシスタントサービスプログラム実行環境３５の音声ＡＩアシスタントサービスプログラムに従って音声ＡＩアシスタントサービスを実行することができる。

[0068] なお、吹き出し５１、５２、５３に配置されるサービス対象物毎にユニークな視覚的特徴には、登場人物の名前の文字列の他、アイコン、吹き出しの色、吹き出しのデザインなど、様々な変形が考えられる。

[0069] 以上のように、本実施形態では、動画コンテンツの映像中の音声ＡＩアシスタントサービスのサービス対象物に、これがサービス対象物であることを

ユーザＵに教示する付加画像を生成し、動画コンテンツの映像に重畳して提示するようにしたので、ユーザＵは映像からサービス対象物を一意に判別して、任意のサービス対象物についてのサービス要求を行うことができる。これにより、ユーザＵから音声ＡＩアシスタントサービス端末３にサービス対象物以外の物体に対するサービス要求を与えるといった無駄な操作が回避され、音声ＡＩアシスタントサービスを良好に利用することができる。

[0070] また、サービス対象物が音声ＡＩアシスタントサービス端末３での音声認識によって一意に認識され得るように、付加画像にサービス対象物毎にユニークな視覚的特徴が付与されたことによって、ユーザＵは、任意のサービス対象物の特徴を表現する音声を使って、そのサービス対象物についてのサービス要求を音声ＡＩアシスタントサービス端末３に音声で与えることができる。これにより、ユーザＵはサービス対象物をどのように表現して音声ＡＩアシスタントサービスに通知するかを戸惑うことがなくなるとともに、音声ＡＩアシスタントサービス端末３においてサービス対象物が確実に特定されるので、良好な音声ＡＩアシスタントサービスが実現される。

[0071] 次に、ＰＯＩメタデータの構造について説明する。

[0072] （ＰＯＩメタデータの構造）

図６はＰＯＩメタデータの構成を示すブロック図である。

ＰＯＩメタデータは、ＰＯＩアイコンイメージ、ＰＯＩ提示色、ＰＯＩ提示位置、ＰＯＩ提示テキスト、ＰＯＩフィルタリング情報を含む。

[0073] ＰＯＩアイコンイメージは、付加画像にサービス対象部の視覚的特徴としてアイコンを提示する場合に用いられるアイコンファイルの実体あるいはアイコンファイルへの参照ＵＲＬ（Ｕｎｉｆｏｒｍ Ｒｅｓｏｕｒｃｅ Ｌｏｃａｔｏｒ）である。ＰＯＩアイコンイメージのＸＭＬ（Ｅｘｔｅｎｓｉｂｌｅ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）表現を以下に示す。

<POIIcon iconPng='true'>…(base64エンコードされたpngファイル)…</POIIcon>

<POIIcon iconPngURL='true'>http://…(アイコンpngファイルのURL)…<

/POIIcon>

[0074] ＰＯＩ提示色は、付加画像に色による視覚的特徴を与える場合に用いられる。ＰＯＩ提示色の情報には、例えば、ＣＳＳ（Ｃａｓｃａｄｉｎｇ Ｓｔｙｌｅ Ｓｈｅｅｔｓ）のカラーコードなどが用いられる。ＰＯＩ提示色のＸＭＬ表現を以下に示す。

<POIColor>...(CSS color code)…</POIColor>

[0075] ＰＯＩ提示位置は、付加画像を提示する対象コンテンツのＵＲＩ（Ｕｎｉｆｏｒｍ Ｒｅｓｏｕｒｃｅ Ｉｄｅｎｔｉｆｉｅｒ）、ＰＯＩメタデータに基づき付加画像を提示する処理を行うアプリケーションのＵＲＩ、付加画像の提示時間、付加画像の提示位置などの情報を含む。ＰＯＩ提示位置のＸＭＬ表現を以下に示す。

<POITargetContentID URL='http://a.com/番組A.MPD'/>

<POIApplication URL='http://a.com/番組AのためのPOI提示アプリ.html(POI 提示アプリのURL)'/>

<POITimePosition start='P0Y0M0DT1H15M2.000S(開始時刻)'end='P0Y0M0DT1H15M2.500S(終了時刻)'/>

<POISPosition x='345(x 座標ピクセル)'y='567(y 座標ピクセル)'/>

[0076] ＰＯＩ提示テキストは、付加画像に文字による視覚的特徴を与える場合にために付加画像に提示される文字列である。ＰＯＩ提示テキストのＸＭＬ表現を以下に示す。

<POIText>…(提示文字列)…</POIText>

[0077] ＰＯＩフィルタリング情報は、付加画像のユーザあるいはユーザ属性に応じた提示制限を行う場合に用いられる。ＰＯＩフィルタリング情報は、付加画像を提示する対象のユーザＵを特定するための情報であり、ユーザデモグラフィッククラスの識別名と任意のユーザ識別情報を含む。

[0078] ユーザデモグラフィッククラスとは、ユーザの性別、年齢、住んでいる地域、職業、学歴、家族構成などの属性をもとにユーザＵを分類するクラスのことであり、ユーザデモグラフィッククラスの識別名によって一意に識別さ

れる。ユーザデモグラフィッククラスの識別名のＸＭＬ表現を以下に示す。

<POITargetUser demographicClass='true'>…(ユーザデモグラフィッククラス識別名)…</POITargetUser>

[0079] 任意のユーザ識別情報は、例えば、放送局関連サービスのアカウント名などのユーザ識別情報である。任意のユーザ識別情報のＸＭＬ表現を以下に示す。

<POITargetUser privateAccount='true'>https://…(サービスのユーザアカウント識別URL等)…</POITargetUser>

[0080] なお、ＰＯＩアイコンイメージ、ＰＯＩ提示色およびＰＯＩ提示テキストは、少なくともいずれか１つが定義されればよい。ＰＯＩフィルタリング情報は、後述する付加情報の提示制限を導入する場合に定義される。

[0081] （付加画像の提示制限について）

上記の実施形態では、動画コンテンツの映像中の音声ＡＩアシスタントサービスのサービス対象物に付加画像を提示することとした。しかしながら、例えば、図７に示すように、一つのシーンに多数のサービス対象物Ｊ１－Ｊ５が存在する場合にはそれらのサービス対象物Ｊ１－Ｊ５の数分の付加画像５１－５５が提示されるため、これらの付加画像５１－５５によって番組の映像の一部が隠れてしまい、番組映像の見た目が損なわれるおそれがある。

[0082] そこで、例えば図８に示すように、ユーザＵに応じて付加画像を提示するサービス対象物を制限することが有効である。例えば、ユーザＵにとって興味のあるサービス対象物に対しては付加画像を提示し、そうではないサービス対象物に対しては付加画像を提示しないようにする。これにより、付加画像によって番組の映像の一部が多数の付加画像に隠れて映像全体が汚れてしまうことを最小限に抑えることができる。この機能を「付加画像の提示制限」と呼ぶこととする。

[0083] 図９は、この付加画像の提示制限を含む動作の流れを示すシーケンス図である。

ここで、ステップＳ４０１－Ｓ４０５の動作は、図２に示した動作の流れ

（その１）のステップＳ１０１－Ｓ１０５と同じであるから、説明を省略する。

[0084] 付加画像の提示制限が行わるために、情報処理装置４のユーザ判別モジュール４７によってユーザＵのユーザデモグラフィッククラスあるいはユーザ識別情報が判別される（ステップＳ４０６）。判別されたユーザＵのユーザデモグラフィッククラスあるいはユーザ識別情報はアプリケーション実行環境４３に通知される（ステップＳ４０７）。

[0085] ユーザデモグラフィッククラスとは、ユーザの性別、年齢、住んでいる地域、職業、学歴、家族構成などの様々な属性をもとにユーザを分類するクラスである。例えば、２０歳代男性であれば最近人気の出てきた新人女優に興味がある傾向が高いことが統計的に言える場合には、２０歳代男性のユーザデモグラフィッククラスは、番組において新人女優が演じる登場人物（サービス対象物）についてのＰＯＩメタデータに定義されたユーザデモグラフィッククラスと一致する。

[0086] ユーザ判別モジュール４７によるユーザＵのユーザデモグラフィッククラスあるいはユーザ識別情報の判別方法には次のようなものが挙げられる。

１．ユーザ判別モジュール４７は、カメラで撮像したユーザＵの顔画像の解析結果からユーザＵの性別や年齢層などの属性を推定し、推定した属性からユーザデモグラフィッククラスを判別する。

２．ユーザ判別モジュール４７は、音声ＡＩアシスタントサービス端末３からユーザＵへの音声による質問を通して得られた情報を基にユーザＵの属性を推定し、ユーザデモグラフィッククラスを判別する。

３．情報処理装置４を使用する複数のユーザが限定されている場合において、各ユーザのユーザ名に対応付けてユーザデモグラフィッククラスの識別名あるいはユーザ識別情報を予め登録しておくことで、ユーザ判別モジュール４７は、生体認証、カード認証などの認証を通して確認されたユーザ名から対応するユーザデモグラフィッククラスの識別名あるいはユーザ識別情報を判別することができる。

[0087] 次に、情報処理装置４のアプリケーション実行環境４３は、動画コンテンツの各シーンの映像に対するすべてのＰＯＩメタデータの中から、ユーザ判別モジュール４７によって判別されたユーザデモグラフィッククラスの識別名あるいはユーザ識別情報がＰＯＩフィルタリング情報として定義されたＰＯＩメタデータを抽出し、抽出したＰＯＩメタデータに基づいて映像中のサービス対象物をユーザＵに教示する付加画像を生成し、レンダラ４４に供給する。これにより、番組の映像の上に付加画像が重畳された合成像が得られ、ディスプレイ４５に表示される（ステップＳ４０８）。

[0088] その後、ユーザＵは、ディスプレイ４５に表示された映像中で付加画像が提示されたサービス対象物についての音声によるサービス要求を音声ＡＩアシスタントサービス端末３に与えて、音声ＡＩアシスタントサービスを利用する（ステップＳ４０９）。

[0089] なお、ここでは、図２に示した情報処理システムの全体的動作の流れ（その１）を前提した付加画像の提示制限を説明したが、全体的動作の流れ（その２）および全体的動作の流れ（その３）において同様である。

[0090] 次に、図８に戻って説明を続ける。ここで、映像中のサービス対象物Ｊ１のユーザデモグラフィッククラスの識別名は”ｃｌａｓｓ１”とし、その他のサービス対象物Ｊ２－Ｊ５のユーザデモグラフィッククラスの識別名は”ｃｌａｓｓ１”以外であったとする。一方、ユーザ判別モジュール４７によってユーザＵのユーザデモグラフィッククラスが”ｃｌａｓｓ１”と判別されたとする。この場合、サービス対象物Ｊ１に対してだけ付加画像５１が提示されることによって、番組の映像の一部分がユーザＵにとって興味のないサービス対象物Ｊ２－Ｊ５に付加される付加画像５２－５５に隠れて映像全体が汚れることを最小限に抑えることができる。

[0091] なお、付加画像の提示制限がユーザ識別情報をフィルタリング条件として行われる場合には、特定のユーザＵだけに特定のサービス対象物に対する付加画像を提示できるように
することができる。

[0092] （ユーザＵの視聴履歴に基づくユーザデモグラフィッククラスの設定）

なお、上記の説明では、性別、年齢、住んでいる地域、職業、学歴、家族内属性などの属性からユーザＵのユーザデモグラフィッククラスを判別することとしたが、ユーザ判別モジュール４７にて、ユーザＵの視聴履歴をもとにユーザＵの嗜好的な条件を算出し、この嗜好的な条件をもとに、あるいは、この嗜好的な条件を加味して、ユーザデモグラフィッククラスを判別してもよい。

[0093] （ＰＯＩメタデータに基づくトリックプレイ再生）

次に、ＰＯＩメタデータに基づくトリックプレイ再生について説明する。

ＰＯＩメタデータに基づくトリックプレイ再生とは、ユーザＵのユーザデモグラフィッククラスあるいはユーザ識別情報を基に抽出されたＰＯＩメタデータに基づいて付加画像が提示されるシーンについては第１の倍速での再生を行い、その他のシーンについては第１の倍速よりも高速な第２の倍速で再生することを言う。

[0094] ここで、第１の倍速は、例えば１倍速（等倍速）あるいは１倍速より低い倍速である。第２の倍速は例えば１倍速よりも高速な早送り再生である。

[0095] 次に、このＰＯＩメタデータに基づくトリックプレイ再生の動作をより詳しく説明する。

まず、ユーザ判別モジュール４７によってユーザＵのユーザデモグラフィッククラスあるいはユーザ識別情報を判別し、アプリケーション実行環境４３に供給する。

[0096] アプリケーション実行環境４３は、番組の各シーンの映像に対するすべてのＰＯＩメタデータの中から、ユーザ判別モジュール４７によって判別されたユーザデモグラフィッククラスの識別名あるいはユーザ識別情報がＰＯＩフィルタリング情報として定義されたＰＯＩメタデータを選択し、抽出したＰＯＩメタデータに基づいて、トリックプレイ再生を行う。

[0097] 図１０はＰＯＩメタデータに基づくトリックプレイ再生をより具体的に示す図である。

ここで、サービス対象物である”Ｂｉｌｌ”と”Ｓａｍ”のユーザデモグラフィッククラスは”ｃｌａｓｓ１”であり、ユーザ判別モジュール４７によって判別されたユーザＵのユーザデモグラフィッククラスが”ｃｌａｓｓ１”であったとする。

[0098] Ｔａ－Ｔａ’期間の映像にはユーザデモグラフィッククラスが”ｃｌａｓｓ１”である”Ｂｉｌｌ”が登場しているため、アプリケーション実行環境４３は、このＴａ－Ｔａ’期間、第１の倍速で付加画像を含む映像の再生を行う。その後時刻Ｔｃになるまで、ユーザデモグラフィッククラスが”ｃｌａｓｓ１”であるサービス対象物が映像に登場しないため、アプリケーション実行環境４３は、第１の倍速よりも高速な第２の倍速で再生を行う。なお、第２の倍速での再生時は付加画像を提示しないことによってアプリケーション実行環境４３の負担を減らすことができる。Ｔｃ－Ｔｃ’期間の映像にはユーザデモグラフィッククラスが”ｃｌａｓｓ１”である”Ｓａｍ”が登場するため、アプリケーション実行環境４３は、このＴｃ－Ｔｃ’期間、第１の倍速で付加画像を含む映像の再生を行う。時刻Ｔｃ’以後はユーザデモグラフィッククラスが”ｃｌａｓｓ１”であるサービス対象物が映像に登場しないため第２の倍速での再生が行われる。

[0099] このように、ユーザＵのユーザデモグラフィッククラスあるいはユーザ識別情報を基に抽出されたＰＯＩメタデータに基づいて付加画像が提示されるシーンについては第１の倍速での再生を行い、その他のシーンについては第１の倍速よりも高速な第２の倍速で再生することによって、ユーザＵにとって有益な（興味のある）シーンにフォーカスしたトリックプレイ再生が実現される。

[0100] また、ユーザＵのユーザデモグラフィッククラスあるいはユーザ識別情報を基に抽出されたＰＯＩメタデータに基づいて付加画像が提示されるシーンだけをスキップ再生してもよい。

[0101] ここまで本実施形態の情報処理システムによるＰＯＩメタデータに基づく付加画像の提示、付加画像の提示制限、トリックプレイ再生等の機能につい

て説明した。

[0102] （アプリケーション実行環境４３について）

図１１はＰＯＩメタデータを処理するアプリケーション実行環境４３の例を示す図である。

本例では、ＰＯＩメタデータを処理するアプリケーションが、情報処理装置４のネイティブのオペレーティングシステム４８の上で動作するネイティブアプリケーション４９である場合を示している。

[0103] 図１２はＰＯＩメタデータを処理するアプリケーション実行環境４３のその他の例を示す図である。

この例では、ＰＯＩメタデータを処理するアプリケーションが、ウェブブラウザ５６上で動作するウェブアプリケーション５７である場合を示している。この場合、ＰＯＩメタデータとウェブアプリケーション５７とが互いに同時あるいは略同時に情報処理装置４に配信される。

[0104] （ＰＯＩメタデータの転送方法）

情報処理装置４にウェブアプリケーションとＰＯＩメタデータとを同時に配信するために、Ｍｕｌｔｉ－ｐａｒｔ ＭＩＭＥ（Ｍｕｌｔｉｐｕｒｐｏｓｅ Ｉｎｔｅｒｎｅｔ ＭＡｉｌ Ｅｘｔｅｎｓｉｏｎｓ）フォーマットを用いて両者をパッケージングする方法がある。図１３はこのウェブアプリケーションとＰＯＩメタデータをパッケージングするＭｕｌｔｉ－ｐａｒｔ ＭＩＭＥフォーマットの例を示す図である。このＭｕｌｔｉ－ｐａｒｔ ＭＩＭＥフォーマットでは、ｂｏｕｎｄａｒｙ－ｐａｒｔによって区切られた各部分にＰＯＩメタデータのファイル６１、ウェブアプリケーションのファイル６２がそれぞれ個別に格納される。

[0105] なお、ＰＯＩメタデータを処理するアプリケーションが、オペレーティングシステムをアプリケーション実行環境として動作するネイティブアプリケーションである場合や、ＰＯＩメタデータとは別途配信されるものである場合には、Ｍｕｌｔｉ－ｐａｒｔ ＭＩＭＥフォーマットにＰＯＩメタデータファイルのみを格納して配信してもよい。

[0106] 次に、ＭＰ４ファイルのフォーマットに映像データおよび音声データなどのメディアデータのトラックと同様にＰＯＩメタデータのトラックを格納して転送する方式を説明する。

[0107] 図１４はＭＰ４ファイルのフォーマットにおけるＭｅｄｉａ　Ｓｅｇｍｅｎｔの構成を示す図である。

同図に示すように、Ｍｅｄｉａ　Ｓｅｇｍｅｎｔは複数のＭｏｖｉｅ　Ｆｒａｇｅｎｔを有し、各々のＭｏｖｉｅ　Ｆｒａｇｅｎｔはｍｏｏｆボックスとｍｄａｔボックスで構成される。ｍｄａｔボックスには、メディアデータが例えばフレームなどの時間の単位で複数のＳａｍｐｌｅボックスに分割されてランダムアクセス可能に格納される。ｍｏｏｆボックスには、ｍｄａｔボックスの各Ｓａｍｐｌｅボックスのメディアデータを提示するタイミングを生成するための情報など、提示に関するメタデータが格納される。

[0108] 本実施形態では、ｍｄａｔボックスの各Ｓａｍｐｌｅボックスに映像データが格納されたＭｅｄｉａ　Ｓｅｇｍｅｎｔ、ｍｄａｔボックスの各Ｓａｍｐｌｅボックスに音声データが格納されたＭｅｄｉａ　Ｓｅｇｍｅｎｔ、ｍｄａｔボックスの各Ｓａｍｐｌｅボックスにメタデータが格納されたＭｅｄｉａ　Ｓｅｇｍｅｎｔが用意される。

[0109] なお、本実施形態では、ＭＰ４ファイルがＭＰＥＧ－ＤＡＳＨ（Ｄｙｎａｍｉｃ　Ａｄａｐｔｉｖｅ　Ｓｔｒｅａｍｉｎｇ　ｏｖｅｒ　ＨＴＴＰ）のＭｅｄｉａ　Ｓｅｇｍｅｎｔであることを前提とする。

[0110] （ＭＰＥＧ－ＤＡＳＨについて）

ＭＰＥＧ－ＤＡＳＨでは、ストリーミング再生が途切れないように、１つの動画コンテンツについて符号化速度と画面サイズが異なる複数のデータ群が用意される。これら複数のデータ群は、情報処理装置４の画面サイズやネットワーク帯域の状態などを考慮して動的に選択される。そのためＭＰＥＧ－ＤＡＳＨでは、上記のように１つの動画コンテンツについてどのような符号化速度と画面サイズのデータ群が用意されているかがＭＰＤ（Ｍｅｄｉａ　Ｐｒｅｓｅｎｔａｔｉｏｎ　Ｄｅｓｃｒｉｐｔｉｏｎ）と呼ばれるメタデータ

に記述される。

[0111] 次に、ＭＰＤのデータ構造の詳細を説明する。

ＭＰＤはサーバに格納された動画コンテンツの構成に関する情報をＸＭＬ（ｅｘｔｅｎｓｉｂｌｅ ｍａｒｋｕｐ ｌａｎｇｕａｇｅ）形式の階層構造で記述したものである。情報処理装置４は、目的の動画コンテンツに対応するＭＰＤファイルをＭＰＤファイルサーバから取得し、解析することによって、サーバから目的の動画コンテンツの提示に必要なＭｅｄｉａ Ｓｅｇｍｅｎｔを取得する。

[0112] 図１５はＭＰＤのデータ構造を示す図である。

ＭＰＤは、その下に１つのＰｅｒｉｏｄと、その下位に各メディアのタイプごとに一つずつのＡｄａｐｔａｔｉｏｎＳｅｔと、さらにその下位の複数のＲｅｐｒｅｓｅｎｔａｔｉｏｎとを有する。

[0113] ＭＰＤの階層つまり最上位階層には、１本の動画コンテンツに関する管理情報として、ＭＰＤの起点、タイトル、ストリーミング種別（オンデマンド／ライブ配信）、長さなどの情報を含む。

[0114] Ｐｅｒｉｏｄは１つの動画コンテンツを、フレームなどの時間で分割した単位である。Ｐｅｒｉｏｄには開始時刻（ｓｔａｒｔ時刻）と終了時刻（ｅｎｄ時刻）が定義される。Ｐｅｒｉｏｄは複数のＡｄａｐｔａｔｉｏｎＳｅｔで構成される。

[0115] ＡｄａｐｔａｔｉｏｎＳｅｔは、Ｐｅｒｉｏｄ毎の動画コンテンツのメディアタイプ（映像、音声、字幕、ＰＯＩメタデータ）ごとのデータに関するコーデック情報、言語などの情報を含む。ＡｄａｐｔａｔｉｏｎＳｅｔは、その下位に符号化速度や画像サイズの異なるデータごとのＲｅｐｒｅｓｅｎｔａｔｉｏｎを有する。

[0116] Ｒｅｐｒｅｓｅｎｔａｔｉｏｎは、それぞれウェブサーバに蓄積される、符号化速度や画像サイズなどの異なるセグメントの符号化速度、画像サイズ、置き場所（ＵＲＬ）などの情報を含む。

[0117] 図１６はＭＰＥＧ－ＤＡＳＨサーバ１５と情報処理装置４との間でのネッ

トワーク通信によるやりとりを示す図である。

ＭＰＥＧ－ＤＡＳＨサーバ１５には、ＭＰＤファイル、および動画コンテンツの各種メディアのＭｅｄｉａ　Ｓｅｇｍｅｎｔが格納される。

[0118] 情報処理装置４のＣＰＵは、ＭＰＥＧ－ＤＡＳＨサーバ１５にＭＰＤファイルを要求する（ステップＳ５０１）。ＭＰＥＧ－ＤＡＳＨサーバ１５は、この要求に対してＭＰＤファイルを情報処理装置４に送信する（ステップＳ５０２）。情報処理装置４のＣＰＵは、受信したＭＰＤファイルを解析して、どのような符号化速度と画像サイズのＭｅｄｉａ　Ｓｅｇｍｅｎｔが用意されているかを確認する（ステップＳ５０３）。情報処理装置４のＣＰＵは、このＭＰＤファイルの解析結果とディスプレイの画面サイズや伝送路のネットワークトラフィクの状態などを考慮して最適な画像サイズや符号化速度のＭｅｄｉａ　Ｓｅｇｍｅｎｔを ＭＰＥＧ－ＤＡＳＨサーバ１５に要求する（ステップＳ５０４）。ＭＰＥＧ－ＤＡＳＨサーバ１５は、この要求に対してＭｅｄｉａ　Ｓｅｇｍｅｎｔを情報処理装置４に送信する（ステップＳ５０５）。

[0119] 次に、上記ＭＰＤとＭｅｄｉａ　Ｓｅｇｍｅｎｔとの関係を説明するために、図１４に戻って、Ｍｅｄｉａ　Ｓｅｇｍｅｎｔにおけるｍｏｏｆボックスとｍｄａｔボックスのデータ構造を説明する。

[0120] 前述したように、Ｍｅｄｉａ　Ｓｅｇｍｅｎｔは複数のＭｏｖｉｅ　Ｆｒａｇｅｎｔを有し、各々のＭｏｖｉｅ　Ｆｒａｇｅｎｔはｍｏｏｆボックスとｍｄａｔボックスで構成される。ｍｄａｔボックスには、メディアデータが例えばフレームなどの時間の単位で複数のＳａｍｐｌｅボックスに分割されてランダムアクセス可能に格納される。ｍｏｏｆボックスには、ｍｄａｔボックスの各Ｓａｍｐｌｅのメディアデータを提示するタイミングを生成するための情報など、提示に関するメタデータが格納される。

[0121] それぞれのＳａｍｐｌｅ（１），（２），（３），…のメディアデータの提示開始時刻ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ（１），（２），（３），…を生成するための情報として、ｍｏｏｆボックスの所定の下位ボックス（

ｍｏｏｆ／ｔｒａｆボックス）にはＢａｓｅＭｅｄｉａＤｅｃｏｄｅＴｉｍｅが格納され、さらにその他のボックス（ｔｒｕｎボックス）には、Ｓａｍｐｌｅ（１），（２），（３），…にそれぞれ対応付けられた情報として、ＳａｍｐｌｅＣｏｕｎｔ（１），（２），（３），…、ＳａｍｐｌｅＤｕｒａｔｉｏｎ（１），（２），（３），…、ＣｏｍｐｏｓｉｔｉｏｎＴｉｍｅＯｆｆｓｅｔ（１），（２），（３），…が格納される。ＢａｓｅＭｅｄｉａＤｅｃｏｄｅＴｉｍｅは、Ｐｅｒｉｏｄの起点からＭｏｖｉｅ　Ｆｒａｇｍｅｎｔの起点までの相対時間の情報である。ＳａｍｐｌｅＣｏｕｎｔ（１），（２），（３），…はＳａｍｐｌｅの数、ＳａｍｐｌｅＤｕｒａｔｉｏｎ（１），（２），（３），…はＳａｍｐｌｅ（１），（２），（３），…の長さ、ＣｏｍｐｏｓｉｔｉｏｎＴｉｍｅＯｆｆｓｅｔ（１），（２），（３），…は調整時間である。

[0122] 次に、Ｓａｍｐｌｅの提示開始時刻ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅの計算方法を示す。

Ｎ番目のＳａｍｐｌｅの提示開始時刻をＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ（Ｎ）とすると、ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ（Ｎ）は、ＢａｓｅＭｅｄｉａＤｅｃｏｄｅＴｉｍｅ＋（Ｎ－１番目までのＳａｍｐｌｅ（１），…，（Ｎ－１）のＳａｍｐｌｅＤｕｒａｔｉｏｎ（１），…，（Ｎ－１）の合計）＋（Ｎ番目のＳａｍｐｌｅのＣｏｍｐｏｓｉｔｉｏｎＴｉｍｅＯｆｆｓｅｔ）（Ｎ）により算出される。

[0123] （ＭＰＥＧ－ＤＡＳＨ動画コンテンツの典型的な提示制御）

図１７はＭＰＥＧ－ＤＡＳＨ動画コンテンツの提示制御の流れを示す図である。

同図において、横軸は実時間（ＵＴＣ　ｔｉｍｅ）の軸とする。情報処理装置４のＣＰＵは、ＭＰＤファイルに定義されたＭＰＤの起点を基準に、ＰｅｒｉｏｄにＭＰＤの起点に対する相対時間として定義された開始時刻をもとに最初のＰｅｒｉｏｄの実時間上の起点を生成する。

[0124] 続いて、情報処理装置４のＣＰＵは、ＢａｓｅＭｅｄｉａＤｅｃｏｄｅＴ

ｉｍｅに基づいてＭｏｖｉｅ　Ｆｒａｇｍｅｎｔの実時間上の起点を生成し、さらにＳａｍｐｌｅＣｏｕｎｔ、ＳａｍｐｌｅＤｕｒａｔｉｏｎ、ＣｏｍｐｏｓｉｔｉｏｎＴｉｍｅＯｆｆｓｅｔを用いて、Ｐｅｒｉｏｄの最初のＳａｍｐｌｅ（１）の提示開始時刻（ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ（１））を生成し、その時刻から最初のＳａｍｐｌｅ（１）の提示を開始する。続いて、情報処理装置４のＣＰＵは、次のＳａｍｐｌｅ（２）の提示開始時刻（ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ（２））を同様に生成して、その時刻に提示する対象をＳａｍｐｌｅ（１）からＳａｍｐｌｅ（２）に切り替える。以降同様にＳａｍｐｌｅの提示の切り替えが行われる。このようにして、各Ｓａｍｐｌｅ（１），（２），…の映像が時間的に途切れることなく提示される。

[0125] 本実施形態では、ＭＰＤファイルにＰＯＩメタデータに対応するＡｄａｐｔａｔｉｏｎＳｅｔを追加したことによって、ＭＰＥＧ－ＤＡＳＨ動画コンテンツの提示制御方法をそのまま踏襲して付加画像の提示制御を行うことが可能となる。これにより、ＭＰＥＧ－ＤＡＳＨ動画コンテンツの映像および音声などの各メディアと同期させて付加画像を提示させることができる。

[0126] 例えば、図１８に示すように、実時間上Ｔ１からＴ２のＰｅｒｉｏｄ（Ｔ１－Ｔ２）の映像に同期して付加画像を提示する場合には、ＭＰＤのＰｅｒｉｏｄ（Ｔ１－Ｔ２）の下位階層にＰＯＩメタデータのＡｄａｐｔａｔｉｏｎＳｅｔ（Ｔ１－Ｔ２）が追加される。さらに、実時間上Ｔ２からＴ３のＰｅｒｉｏｄ（Ｔ２－Ｔ３）の映像に同期して付加画像を提示する場合には、ＭＰＤのＰｅｒｉｏｄ（Ｔ２－Ｔ３）の下位階層にＰＯＩメタデータのＡｄａｐｔａｔｉｏｎＳｅｔ（Ｔ２－Ｔ３）が追加される。

[0127] 図１９は、ＰＯＩメタデータのＡｄａｐｔａｔｉｏｎＳｅｔが追加されたＭＰＤのより具体的な例を示す図である。

このＭＰＤの最上位階層には@ａｖＡｖａｉｌａｂｉｌｉｔｙＳｔａｒｔＴｉｍｅ=Ｔ０と記述されている。これは、動画コンテンツの時間の起点がＴ０であることを示す。その下位階層には２つのＰｅｒｉｏｄが格納される。２つの

Ｐｅｒｉｏｄのうち、最初のＰｅｒｉｏｄはＴ０の起点から０ｓｅｃ後に開始され、２番目のＰｅｒｉｏｄはＴ０の起点から１００ｓｅｃ後に開始されることが定義される。

[0128] この例では、２番目のＰｅｒｉｏｄのＡｄａｐｔａｔｉｏｎＳｅｔとして映像、音声およびＰＯＩメタデータそれぞれのＡｄａｐｔａｔｉｏｎＳｅｔが存在する。映像のＡｄａｐｔａｔｉｏｎＳｅｔの下位階層には異なる符号化速度（２５６Ｋｂｐｓ／５１２Ｋｂｐｓ）を示した２つのＲｅｐｒｅｓｅｎｔａｔｉｏｎが存在する。それぞれのＲｅｐｒｅｓｅｎｔａｔｉｏｎの下位階層には、映像のＭｅｄｉａ Ｓｅｇｍｅｎｔの置き場所を示すＵＲＬの生成方法や取得周期などが記述されたＳｅｇｍｅｎｔＴｅｍｐｌａｔｅが存在する。

[0129] ここで、”@ｔｉｍｅｓｃａｌｅ＝”１０００””、”@ｄｕｒａｔｉｏｎ＝”１０００””は、１／１０００秒のタイムスケールで１０００分の値つまり１秒をセグメント時間長として、このセグメント時間長の周期で映像を取得できることを意味する。”@Ｍｅｄｉａ=”$Ｎｕｍｂｅｒ％６ｄ$．ｍ４ｓ””は各Ｓａｍｐｌｅの映像の置き場所を示すＵＲＬの最下位の値の生成方法を示し、具体的には６桁の値の１から増加する整数を意味する。例えば、ＳａｍｐｌｅのＵＲＬは、ＭＰＤの各要素に記述された”ＢａｓｅＵＲＬ”が示す値をパス形式に繋ぎ、最後に”@Ｍｅｄｉａ=”$Ｎｕｍｂｅｒ％６ｄ$．ｍ４ｓ””により生成される６桁の値を付加することによって生成される。すなわち、映像の１番目のＳａｍｐｌｅのＵＲＬは”ＨＴＴＰ：／／ａ．ｃｏｍ／ｐ２／ｖｉｄｅｏ／５１２／０００００１．ｍ４ｓ”のように作成される。”@ｉｎｉｔｉａｌｉｚａｔｉｏｎ＝”ＩＳ．ｍｐ４””は、映像の１番目のＭｅｄｉａ ＳｅｇｍｅｎｔのＩｎｉｔｉａｌｉｚａｔｉｏｎ ＳｅｇｍｅｎｔのＵＲＬに用いられる値である。このＩｎｉｔｉａｌｉｚａｔｉｏｎ ＳｅｇｍｅｎｔのＵＲＬについても、ＭＰＤの各要素に記述された”ＢａｓｅＵＲＬ”が示す値をパス形式に繋ぎ、最後に”ＩＳ．ｍｐ４”を付加することによって生成される。例えば、”ＨＴＴＰ：／／ａ．ｃｏｍ／ｐ２／ｖｉｄｅｏ／５１２／

ＩＳ．ｍｐ４”のように作成される。

[0130] ＰＯＩメタデータのＡｄａｐｔａｔｉｏｎＳｅｔの下位階層にはＲｅｐｒｅｓｅｎｔａｔｉｏｎが存在し、さらにその下位階層にはＰＯＩメタデータのＭｅｄｉａ　Ｓｅｇｍｅｎｔの置き場所を示すＵＲＬの生成方法や取得周期などが記述されたＳｅｇｍｅｎｔＴｅｍｐｌａｔｅが存在する。このＰＯＩメタデータのＭｅｄｉａ　Ｓｅｇｍｅｎｔの置き場所を示すＵＲＬの生成方法は映像のＭｅｄｉａ　Ｓｅｇｍｅｎｔの置き場所を示すＵＲＬの生成方法と同じであってよい。また、ＰＯＩメタデータのＩｎｉｔｉａｌｉｚａｔｉｏｎ　Ｓｅｇｍｅｎｔの置き場所を示すＵＲＬの生成方法も映像のＩｎｉｔｉａｌｉｚａｔｉｏｎ　Ｓｅｇｍｅｎｔの置き場所を示すＵＲＬの生成方法と同じであってよい。ＰＯＩメタデータのＩｎｉｔｉａｌｉｚａｔｉｏｎ　Ｓｅｇｍｅｎｔには、Ｍｅｄｉａ　ＳｅｇｍｅｎｔにＰＯＩメタデータがＳａｍｐｌｅとして格納されることを識別する情報が含まれる。具体的には、Ｉｎｉｔｉａｌｉｚａｔｉｏｎ　Ｓｅｇｍｅｎｔのｍｏｏｖ（ＭｏｖｉｅＢｏｘ）／ｔｒａｋ（ＴｒａｃｋＢｏｘ）／ｍｄｉａ（ＭｅｄｉａＢｏｘ）／ｈｄｌｒ（ＨａｎｄｌｅｒＢｏｘ）のｈａｎｄｌｅｒ＿ｔｙｐｅフィールドの値が”ｍｅｔａ”であることによって、ＰＯＩメタデータがＭｅｄｉａ　ＳｅｇｍｅｎｔにＳａｍｐｌｅとして格納されることを識別することができる。

[0131] 情報処理装置４のＣＰＵは、上記のように生成されるＵＲＬに基づいて、動画コンテンツの映像、音声およびＰＯＩメタデータをそれぞれＳａｍｐｌｅの単位で取得することができる。

[0132] 図２０は、ＭＰＤに基づく映像および付加画像の提示の流れを示す図である。

映像の各Ｓａｍｐｌｅ（１），（２），（３）を提示する処理は上述したとおりである。

ここで、情報処理装置４のＣＰＵは、映像の最初のＳａｍｐｌｅ（１）の実時間上の提示開始時刻（ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ）（１）から

次のＳａｍｐｌｅ（２）の実時間上の提示開始時刻（ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ）（２）までの間に、ＰＯＩメタデータ（１）に基づく付加画像の提示処理を行う。この後、情報処理装置４のＣＰＵは、Ｓａｍｐｌｅ（２）の実時間上の提示開始時刻（ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ）（２）からその次のＳａｍｐｌｅ（３）の実時間上の提示開始時刻（ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ）（３）までの間にＰＯＩメタデータ（２）に基づく付加画像の提示処理を行い、さらにＳａｍｐｌｅ（２）の実時間上の提示開始時刻（ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ）（３）からその次のＳａｍｐｌｅ（３）の実時間上の提示開始時刻（ＰｒｅｓｅｎｔａｔｉｏｎＴｉｍｅ）（４）までの間にＰＯＩメタデータ（３）に基づく付加画像の提示処理を行う。

[0133] 以上のように、本実施形態では、ＭＰＤファイルにＰＯＩメタデータのためのＡｄａｐｔａｔｉｏｎＳｅｔを追加することによって、ＰＯＩメタデータに基づく付加画像の提示制御を、映像、音声などの他の種類のメディアの提示制御と同じ仕組みで行うことができるとともに、映像、音声などの他の種類のメディアと付加画像を正確に同期させて提示することができる。

[0134] （付加画像の提示位置の移動）

図２１は映像中のサービス対象物が時間に沿って移動する場合にそのサービス対象物の移動に伴って付加画像の提示位置も移動させる場合のＰＯＩメタデータを示す図である。

ここで、Ｔ１はＰＯＩメタデータ（１）に基づく付加画像の提示開始時刻、Ｔ２はＰＯＩメタデータ（２）に基づく付加画像の提示開始時刻、Ｔ３はＰＯＩメタデータ（３）に基づく付加画像の提示開始時刻である。Ｔ１－Ｔ２はＰＯＩメタデータ（１）に基づく付加画像の提示期間であり、この期間、付加画像は、ＰＯＩメタデータ（１）中のＰＯＩＰｏｓｉｔｉｏｎ要素の値（ｘ＝ｘ１，ｙ＝ｙ１）が示す位置に提示される。Ｔ２－Ｔ３はＰＯＩメタデータ（２）に基づく付加画像の提示期間であり、この期間、付加画像は、ＰＯＩメタデータ（２）中のＰＯＩＰｏｓｉｔｉｏｎ要素の値（ｘ＝ｘ２，ｙ

＝ｙ２）が示す位置に提示される。そしてＴ３－Ｔ４はＰＯＩメタデータ（３）に基づく付加画像の提示期間であり、この期間、付加画像は、ＰＯＩメタデータ（３）中のＰＯＩＰｏｓｉｔｉｏｎ要素の値（ｘ＝ｘ３，ｙ＝ｙ３）が示す位置に提示される。

[0135] このように、各ＳａｍｐｌｅのＰＯＩメタデータにおけるＰＯＩＰｏｓｉｔｉｏｎ要素の値を、移動するサービス対象物の位置に合わせて決めることによって、付加画像の提示位置をサービス対象物の移動に正確に同期させて移動させることができる。

[0136] （ＰＯＩメタデータの複数サンプル時間をまたがる提示更新制御）

ここまで、各Ｓａｍｐｌｅの映像にＰＯＩメタデータを１対１に対応付けて付加画像の提示制御が行われる場合を想定したが、１つのＰＯＩメタデータを連続する複数のＳａｍｐｌｅの映像に適用させてもよい。この場合、図２２に示すように、ＰＯＩメタデータのファイル６３、６４、６５を格納したパッケージ６６、６７、６８に記述されるＰＯＩメタデータの識別子（ｍｅｔａｄａｔａＵＲＩ）の値にバージョン情報（Ｖｅｒｓｉｏｎ）が付加される。このバージョン情報は、直前のＰＯＩメタデータに対して内容の変化がない場合には、直前のＰＯＩメタデータを格納したパッケージに記述されるバージョン情報と同じ値とされ、直前のＰＯＩメタデータに対して内容の変化がある場合にはインクリメントされた値に設定される。

[0137] これにより、アプリケーション実行環境４３のアプリケーションは、映像のＳａｍｐｌｅ間の切り替えにおいて、切り替え前後の各バージョン情報の値が変化した場合には、ＰＯＩメタデータに基づく付加画像の提示のための演算を行い、各バージョン情報の値に変化がない場合には、改めてＰＯＩメタデータに基づく付加画像の提示のための演算を行うことなく付加画像をそのまま提示し続ける。これにより、付加画像の提示のための演算の負荷を低減させることができる。

[0138] 例えば、図２２に示すように、Ｓａｍｐｌｅ（１）の映像に対応するＰＯＩメタデータ（１）の識別子（ｍｅｔａｄａｔａＵＲＩ）に付加されたバー

ジョン情報の値は”１”、２番目のＳａｍｐｌｅ（２）の映像に対応するＰＯＩメタデータ（２）の識別子（ｍｅｔａｄａｔａＵＲＩ）に付加されたバージョン情報の値は”２”、３番目のＳａｍｐｌｅ（３）の映像に対応するＰＯＩメタデータ（３）の識別子（ｍｅｔａｄａｔａＵＲＩ）に付加されたバージョン情報の値は”２”である。この場合、２番目のＳａｍｐｌｅ（２）の映像と３番目のＳａｍｐｌｅ（３）の映像との切り替えにおいて、３番目のＳａｍｐｌｅ（３）の映像に対応するＰＯＩメタデータ（３）に基づく付加画像の提示のための演算を行うことなく、２番目のＳａｍｐｌｅ（２）の映像に付加された付加画像が３番目のＳａｍｐｌｅ（３）の映像にもそのまま続けて提示される。

[0139] なお、本技術は、上述の実施形態にのみ限定されるものではなく、本発明の要旨を逸脱しない範囲内において種々変更を加え得ることは勿論である。

[0140] 本技術は以下のような構成も採ることができる。

（１） ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データを取得して再生するメディア再生部と、

前記再生した映像に前記サービス対象物を前記ユーザに教示するための付加画像を付加する制御部と

を具備する情報処理装置。

[0141] （２）前記（１）に記載の情報処理装置であって、

前記付加画像は、前記サービス対象物が前記サービスにおいて音声認識によって一意に判別され得るように、前記サービス対象物毎にユニークな視覚的特徴を有する

情報処理装置。

[0142] （３）前記（１）から（２）のうちいずれか１つに記載の情報処理装置であって、

前記付加画像は、前記サービス対象物に付随した位置に提示される

情報処理装置。

[0143] （４）前記（１）から（３）のうちいずれか１つに記載の情報処理装置であ

って、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて、前記付加画像が付加される前記サービス対象物を制限するように構成された

情報処理装置。

[0144] （５）前記（１）から（４）のうちいずれか１つに記載の情報処理装置であって、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて前記付加画像が付加された前記サービス対象物が登場する期間の映像をスキップ再生するように構成された

情報処理装置。

[0145] （６）前記（１）から（４）のうちいずれか１つに記載の情報処理装置であって、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて前記付加画像が付加された前記サービス対象物が登場する期間の映像を第１の倍速で再生し、前記サービス対象物が登場しない期間の映像を前記第１の倍速よりも高速な第２の速度で再生するように構成された

情報処理装置。

[0146] （７）前記（１）から（６）のうちいずれか１つに記載の情報処理装置であって、

前記制御部は、前記付加情報を生成するためのメタデータを取得し、前記取得したメタデータに基づき前記付加情報を付加するように構成された

情報処理装置。

[0147] （８）前記（１）から（７）のうちいずれか１つに記載の情報処理装置であって、

前記制御部は、前記メタデータを処理するためのウェブアプリケーションを取得し、前記取得したウェブアプリケーションに従って前記メタデータを処理するように構成された

情報処理装置。

[0148] （９）前記（１）から（８）のうちいずれか１つに記載の情報処理装置であって、

前記制御部は、前記メタデータのＡｄａｐｔａｔｉｏｎＳｅｔを含むＭＰＤファイルを取得し、このＭＰＤファイルを解析して、前記映像データおよび前記メタデータをそれぞれＭＰＥＧ－ＤＡＳＨのＭｅｄｉａ　Ｓｅｇｍｅｎｔとして取得し、前記映像データと、前記メタデータに基づく前記付加画像とを互いに同期させて提示するように構成された

情報処理装置。

[0149] （１０）前記（１）から（９）のうちいずれか１つに記載の情報処理装置であって、

前記制御部は、前記メタデータのバージョン情報に基づいて、時間的に前後する前側のメタデータと後側のメタデータとの内容の変化の有無を判別し、変化がない場合、前記前側のメタデータに基づき映像に付加した付加画像を、前記後側のメタデータに基づく付加画像として、前記後側のメタデータに同期する映像に付加するように構成された

情報処理装置。

[0150] （１１）前記（１）から（１０）のうちいずれか１つに記載の情報処理装置であって、

前記付加画像の視覚的特徴が、前記サービス対象物に関する文字列、色、形状、またはアイコンのいずれか１つによって与えられる

情報処理装置。

[0151] （１２）ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データを取得して再生し、

前記再生した映像に前記サービス対象物を前記ユーザに教示するための付加画像を付加する

情報処理方法。

[0152] （１３）前記（１２）に記載の情報処理方法であって、

前記付加画像は、前記サービス対象物が前記サービスにおいて音声認識に

よって一意に判別され得るように、前記サービス対象物毎にユニークな視覚的特徴を有する

情報処理方法。

[0153] （１４）前記（１２）または（１３）に記載の情報処理方法であって、

前記付加画像は、前記サービス対象物に付随した位置に提示される

情報処理方法。

[0154] （１５）前記（１２）から（１４）のうちいずれか１つに記載の情報処理方法であって、

前記ユーザまたは前記ユーザの属性に応じて、前記付加画像が付加される前記サービス対象物を制限する

情報処理方法。

[0155] （１６）前記（１２）から（１５）のうちいずれか１つに記載の情報処理方法であって、

前記ユーザまたは前記ユーザの属性に応じて前記付加画像が付加された前記サービス対象物が登場する期間の映像をスキップ再生する

情報処理方法。

[0156] （１７）前記（１２）から（１５）のうちいずれか１つに記載の情報処理方法であって、

前記ユーザまたは前記ユーザの属性に応じて前記付加画像が付加された前記サービス対象物が登場する期間の映像を第１の倍速で再生し、前記サービス対象物が登場しない期間の映像を前記第１の倍速よりも高速な第２の速度で再生する

情報処理方法。

[0157] （１８）前記（１２）から（１７）のうちいずれか１つに記載の情報処理方法であって、

前記付加情報を生成するためのメタデータを取得し、前記取得したメタデータに基づき前記付加情報を付加する

情報処理方法。

[0158] （１９）前記（１２）から（１８）のうちいずれか１つに記載の情報処理方法であって、

前記メタデータを処理するためのウェブアプリケーションを取得し、前記取得したウェブアプリケーションに従って前記メタデータを処理する

情報処理方法。

[0159] （２０）前記（１２）から（１９）のうちいずれか１つに記載の情報処理方法であって、

前記メタデータのＡｄａｐｔａｔｉｏｎＳｅｔを含むＭＰＤファイルを取得し、このＭＰＤファイルを解析して、前記映像データおよび前記メタデータをそれぞれＭＰＥＧ－ＤＡＳＨのＭｅｄｉａ　Ｓｅｇｍｅｎｔとして取得し、前記映像データと、前記メタデータに基づく前記付加画像とを互いに同期させて提示する

情報処理方法。

[0160] （２１）前記（１２）から（２０）のうちいずれか１つに記載の情報処理方法であって、

前記メタデータのバージョン情報に基づいて、時間的に前後する前側のメタデータと後側のメタデータとの内容の変化の有無を判別し、変化がない場合、前記前側のメタデータに基づき映像に付加した付加画像を、前記後側のメタデータに基づく付加画像として、前記後側のメタデータに同期する映像に付加する

情報処理方法。

[0161] （２２）前記（１２）から（２１）のうちいずれか１つに記載の情報処理方法であって、

前記付加画像の視覚的特徴が、前記サービス対象物に関する文字列、色、形状、またはアイコンのいずれか１つによって与えられる

情報処理方法。

[0162] （２３）ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データをコンピュータにて再生するとき、前記再生し

た映像に前記サービス対象物を前記ユーザに教示するための付加画像を付加するようにコンピュータを機能させるプログラム。

[0163] （２４）前記（２３）に記載のプログラムであって、

前記付加画像は、前記サービス対象物が前記サービスにおいて音声認識によって一意に判別され得るように、前記サービス対象物毎にユニークな視覚的特徴を有する

プログラム。

[0164] （２５）前記（２３）または（２４）に記載のプログラムであって、

前記付加画像は、前記サービス対象物に付随した位置に提示される

プログラム。

[0165] （２６）前記（２３）から（２５）のうちいずれか１つに記載のプログラムであって、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて、前記付加画像が付加される前記サービス対象物を制限する

プログラム。

[0166] （２７）前記（２３）から（２６）のうちいずれか１つに記載のプログラムであって、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて前記付加画像が付加された前記サービス対象物が登場する期間の映像をスキップ再生する

プログラム。

[0167] （２８）前記（２３）から（２６）のうちいずれか１つに記載のプログラムであって、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて前記付加画像が付加された前記サービス対象物が登場する期間の映像を第１の倍速で再生し、前記サービス対象物が登場しない期間の映像を前記第１の倍速よりも高速な第２の速度で再生する

プログラム。

[0168] （２９）前記（２３）から（２８）のうちいずれか１つに記載のプログラム

であって、

前記制御部は、前記付加情報を生成するためのメタデータを取得し、前記取得したメタデータに基づき前記付加情報を付加する

プログラム。

[0169] （３０）前記（２３）から（２８）のうちいずれか１つに記載のプログラムであって、

前記制御部は、前記メタデータのバージョン情報に基づいて、時間的に前後する前側のメタデータと後側のメタデータとの内容の変化の有無を判別し、変化がない場合、前記前側のメタデータに基づき映像に付加した付加画像を、前記後側のメタデータに基づく付加画像として、前記後側のメタデータに同期する映像に付加する

プログラム。

[0170] （３１）前記（２３）から（２９）のうちいずれか１つに記載のプログラムであって、

前記付加画像の視覚的特徴が、前記サービス対象物に関する文字列、色、形状、またはアイコンのいずれか１つによって与えられる

プログラム。

符号の説明

[0171] ４…情報処理装置

１１…動画コンテンツサーバ

１２…アプリケーションサーバ

１３…ＰＯＩメタデータサーバ

４１…ＡＶデコーダ

４２…ＰＯＩメタデータ処理モジュール

４３…アプリケーション実行環境

４４…レンダラ

４５…ディスプレイ

４６…スピーカ

４７…ユーザ判別モジュール

請求の範囲

[請求項1] ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データを取得して再生するメディア再生部と、

前記再生した映像に前記サービス対象物を前記ユーザに教示するための付加画像を付加する制御部と

を具備する情報処理装置。

[請求項2] 請求項１に記載の情報処理装置であって、

前記付加画像は、前記サービス対象物が前記サービスにおいて音声認識によって一意に判別され得るように、前記サービス対象物毎にユニークな視覚的特徴を有する

情報処理装置。

[請求項3] 請求項２に記載の情報処理装置であって、

前記付加画像は、前記サービス対象物に付随した位置に提示される

情報処理装置。

[請求項4] 請求項３に記載の情報処理装置であって、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて、前記付加画像が付加される前記サービス対象物を制限するように構成された

情報処理装置。

[請求項5] 請求項４に記載の情報処理装置であって、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて前記付加画像が付加された前記サービス対象物が登場する期間の映像をスキップ再生するように構成された

情報処理装置。

[請求項6] 請求項５に記載の情報処理装置であって、

前記制御部は、前記ユーザまたは前記ユーザの属性に応じて前記付加画像が付加された前記サービス対象物が登場する期間の映像を第１の倍速で再生し、前記サービス対象物が登場しない期間の映像を前記

第１の倍速よりも高速な第２の速度で再生するように構成された

情報処理装置。

[請求項7] 請求項６に記載の情報処理装置であって、

前記制御部は、前記付加情報を生成するためのメタデータを取得し、前記取得したメタデータに基づき前記付加情報を付加するように構成された

情報処理装置。

[請求項8] 請求項７に記載の情報処理装置であって、

前記制御部は、前記メタデータを処理するためのウェブアプリケーションを取得し、前記取得したウェブアプリケーションに従って前記メタデータを処理するように構成された

情報処理装置。

[請求項9] 請求項８に記載の情報処理装置であって、

前記制御部は、前記メタデータのＡｄａｐｔａｔｉｏｎＳｅｔを含むＭＰＤファイルを取得し、このＭＰＤファイルを解析して、前記映像データおよび前記メタデータをそれぞれＭＰＥＧ－ＤＡＳＨのＭｅｄｉａ　Ｓｅｇｍｅｎｔとして取得し、前記映像データと、前記メタデータに基づく前記付加画像とを互いに同期させて提示するように構成された

情報処理装置。

[請求項10] 請求項９に記載の情報処理装置であって、

前記制御部は、前記メタデータのバージョン情報に基づいて、時間的に前後する前側のメタデータと後側のメタデータとの内容の変化の有無を判別し、変化がない場合、前記前側のメタデータに基づき映像に付加した付加画像を、前記後側のメタデータに基づく付加画像として、前記後側のメタデータに同期する映像に付加するように構成された

情報処理装置。

[請求項11] 請求項２に記載の情報処理装置であって、

前記付加画像の視覚的特徴が、前記サービス対象物に関する文字列、色、形状、またはアイコンのいずれか１つによって与えられる

情報処理装置。

[請求項12] ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データを取得して再生し、

前記再生した映像に前記サービス対象物を前記ユーザに教示するための付加画像を付加する

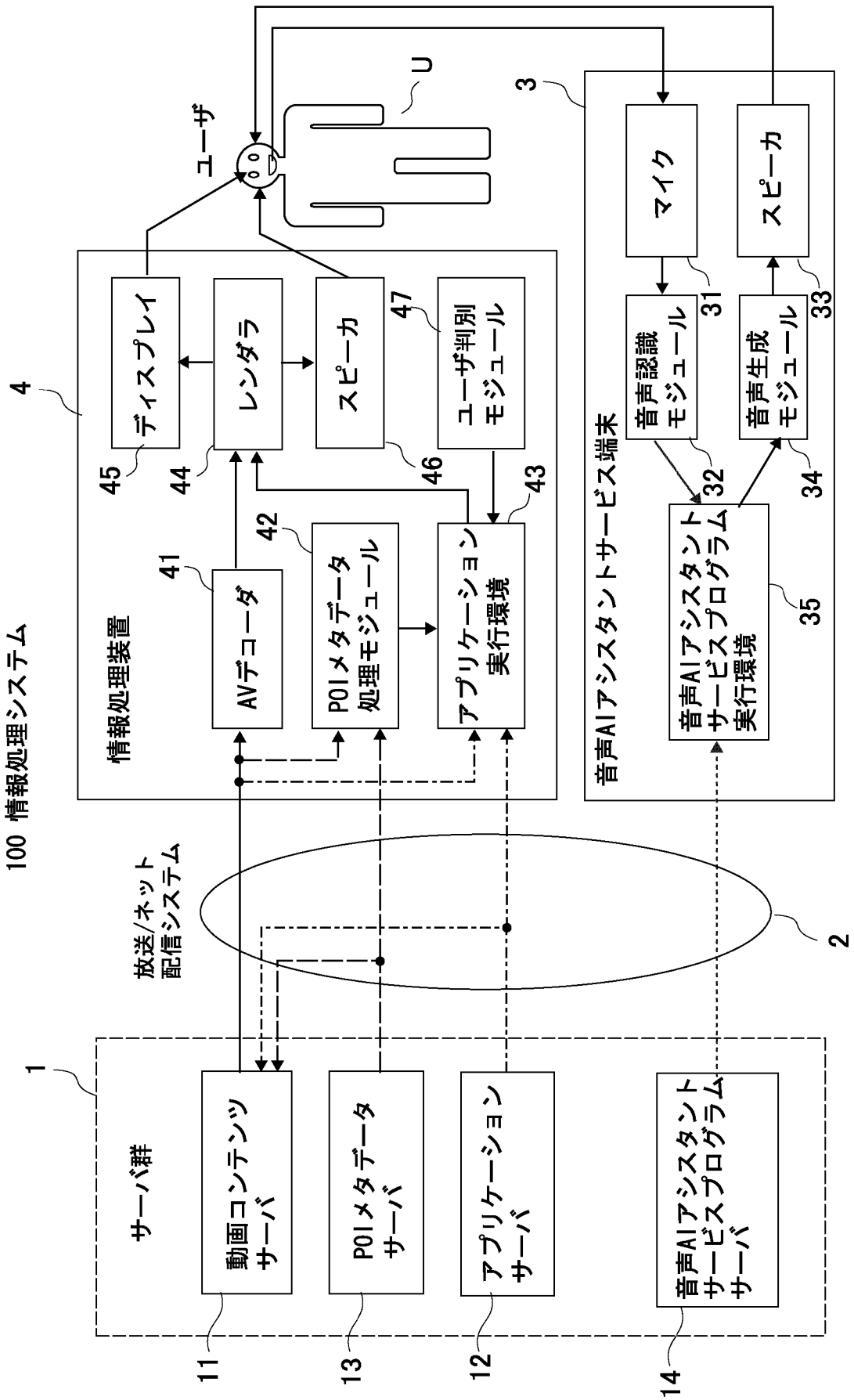
情報処理方法。

[請求項13] ユーザからの音声による要求を処理するサービスを利用可能なサービス対象物を含む映像データを取得して再生するメディア再生部と、

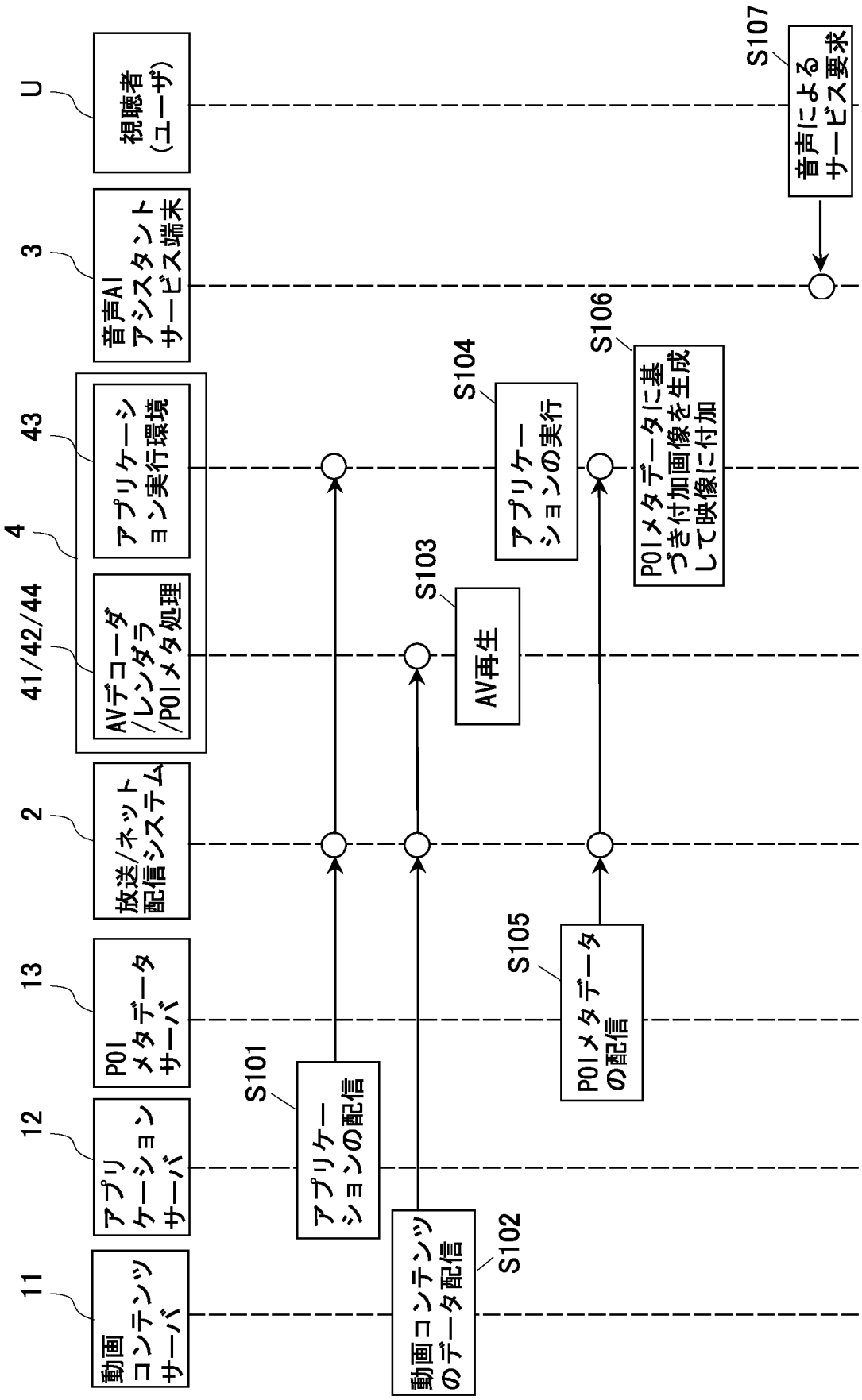
前記再生した映像に前記サービス対象物を前記ユーザに教示するための付加画像を付加する制御部として、

コンピュータを機能させるプログラム。

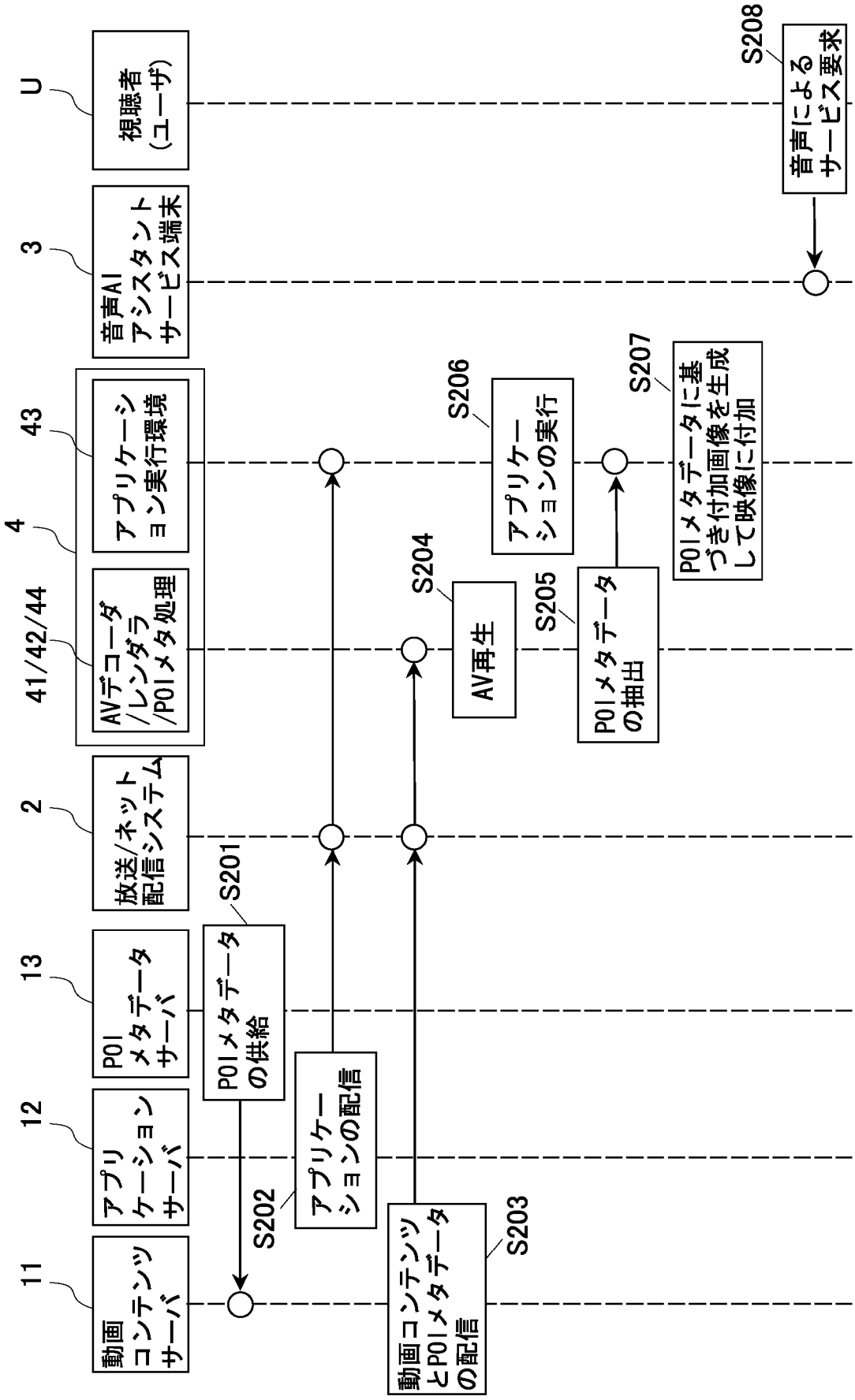
[図1]


100 情報処理システム
ユーザ
U
3
マイク
スピーカ
音声AIアシスタントサービス端末
音声認識モジュール
音声生成モジュール
音声AIアシスタントサービスプログラム実行環境
31
33
32
34
35
4
情報処理装置
ディスプレイ
レンダラ
スピーカ
ユーザ判別モジュール
AVデコーダ
POIメタデータ処理モジュール
アプリケーション実行環境
45
44
46
47
43
41
42
放送/ネット配信システム
2
1
サーバ群
動画コンテンツサーバ
POIメタデータサーバ
アプリケーションサーバ
音声AIアシスタントサービスプログラムサーバ
11
13
12
14

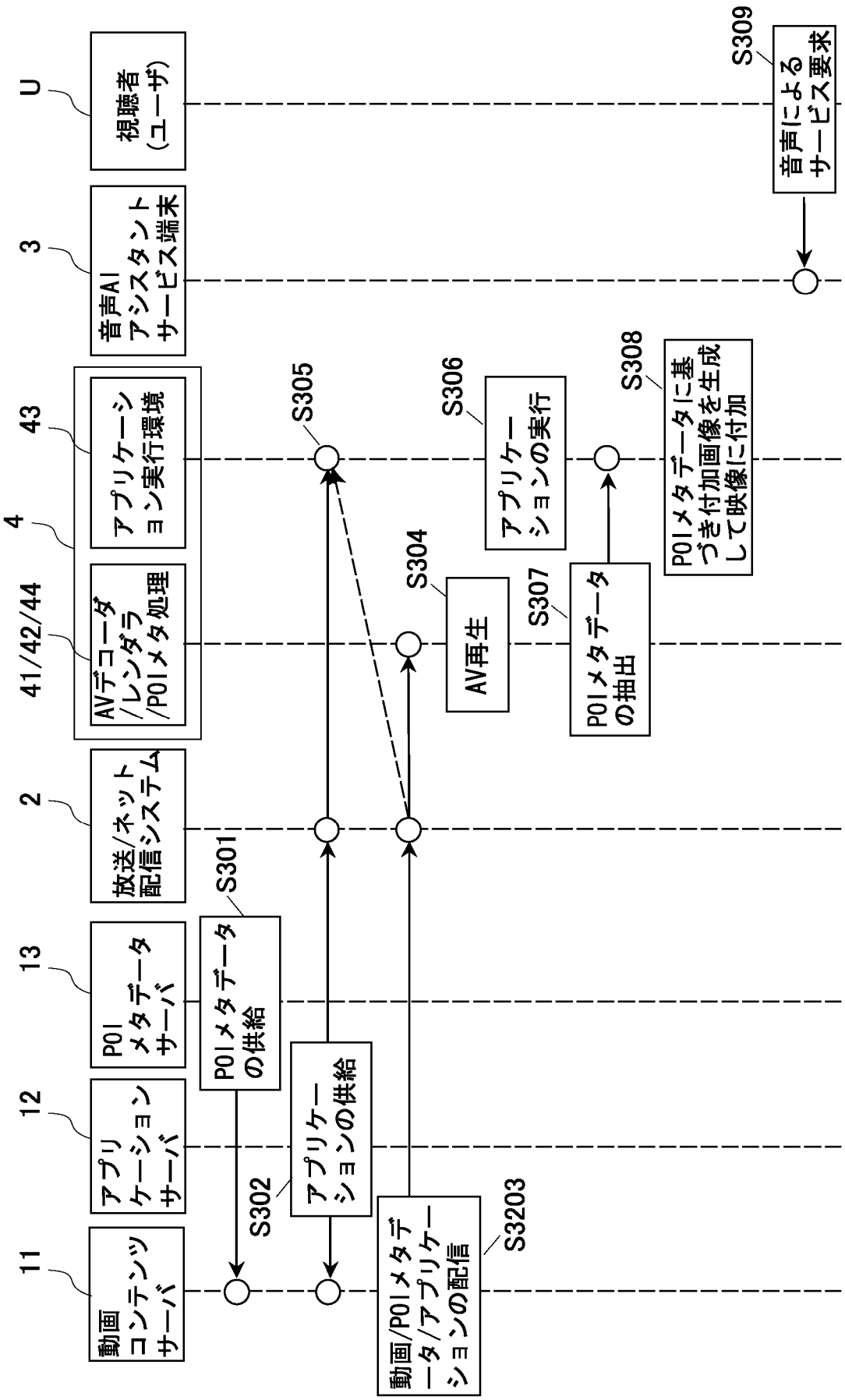
[図2]


U
視聴者
(ユーザ)
3
音声AI
アシスタント
サービス端末
4
43
アプリケーション実行環境
41/42/44
AVデコーダ
/レンダラ
/POIメタ処理
2
放送/ネット
配信システム
13
POI
メタデータ
サーバ
12
アプリ
ケーション
サーバ
11
動画
コンテンツ
サーバ
S101
アプリケーション
の配信
S102
動画コンテンツ
のデータ配信
S103
AV再生
S104
アプリケーション
の実行
S105
POIメタデータ
の配信
S106
POIメタデータに基
づき付加画像を生成
して映像に付加
S107
音声による
サービス要求

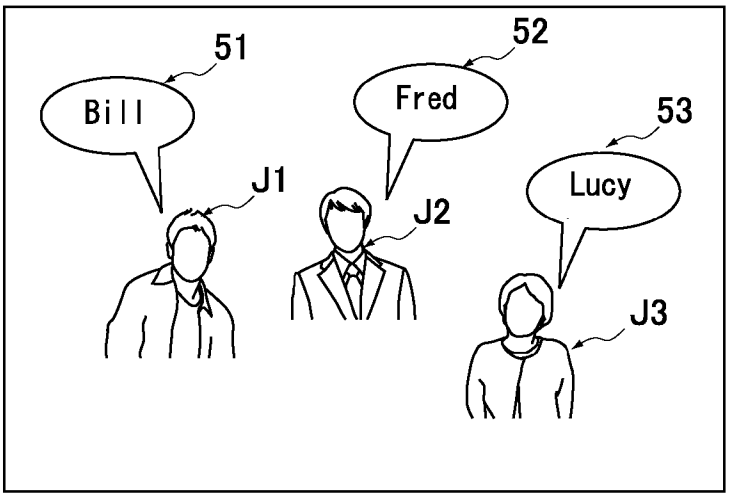
[図3]


U
視聴者
(ユーザ)
3
音声AI
アシスタント
サービス端末
S208
音声による
サービス要求
43
4
アプリケーション実行環境
41/42/44
AVデコーダ
/レンダラ
/POIメタ処理
S206
アプリケーション実行
S207
POIメタデータに基づき付加画像を生成して映像に付加
S204
AV再生
S205
POIメタデータの抽出
2
放送/ネット
配信システム
13
POI
メタデータ
サーバ
S201
POIメタデータの供給
12
アプリ
ケーション
サーバ
アプリケーションの配信
11
動画
コンテンツ
サーバ
S202
動画コンテンツとPOIメタデータの配信
S203

[図4]


U
視聴者
(ユーザ)
3
音声AI
アシスタント
サービス端末
S309
音声による
サービス要求
43
4
アプリケーション実行環境
41/42/44
AVデコーダ
/レンダラ
/POIメタ処理
S305
S306
アプリケーション実行
S308
POIメタデータに基づき付加画像を生成して映像に付加
S304
AV再生
S307
POIメタデータの抽出
2
放送/ネット
配信システム
S301
13
POI
メタデータ
サーバ
POIメタデータの供給
12
アプリ
ケーション
サーバ
アプリケーションの供給
S302
11
動画
コンテンツ
サーバ
動画/POIメタデータ/アプリケーションの配信
S3203

[図5]

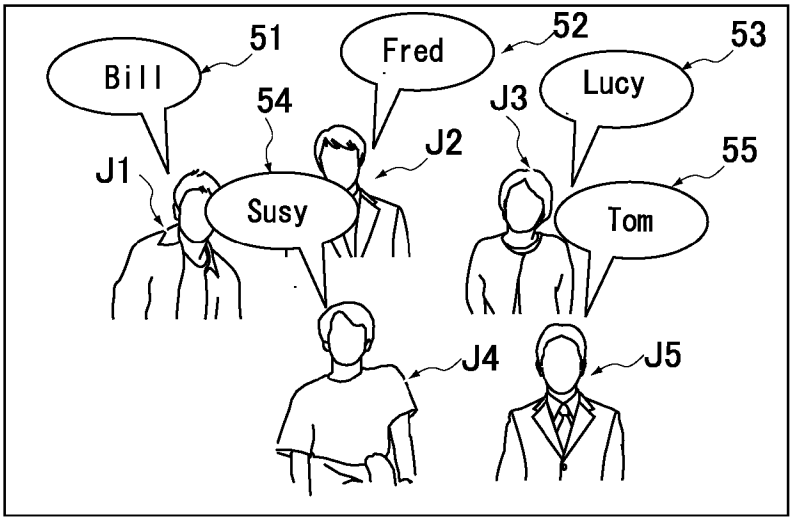


[図6]

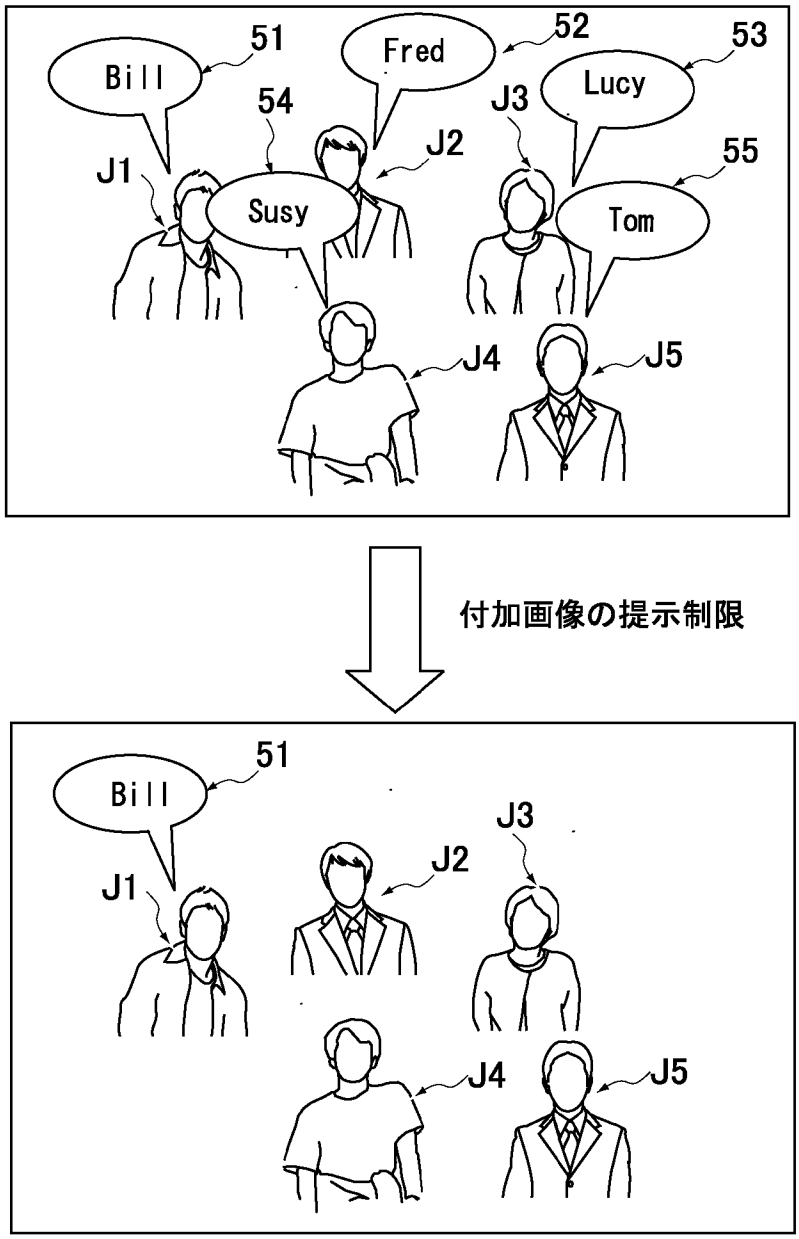
ＰＯＩメタデータ

・ＰＯＩアイコンイメージ
・ＰＯＩ提示色
・ＰＯＩ提示位置
・ＰＯＩ提示テキスト
・ＰＯＩフィルタリング情報
　　ユーザデモグラフィッククラスの識別名
　　ユーザ識別情報

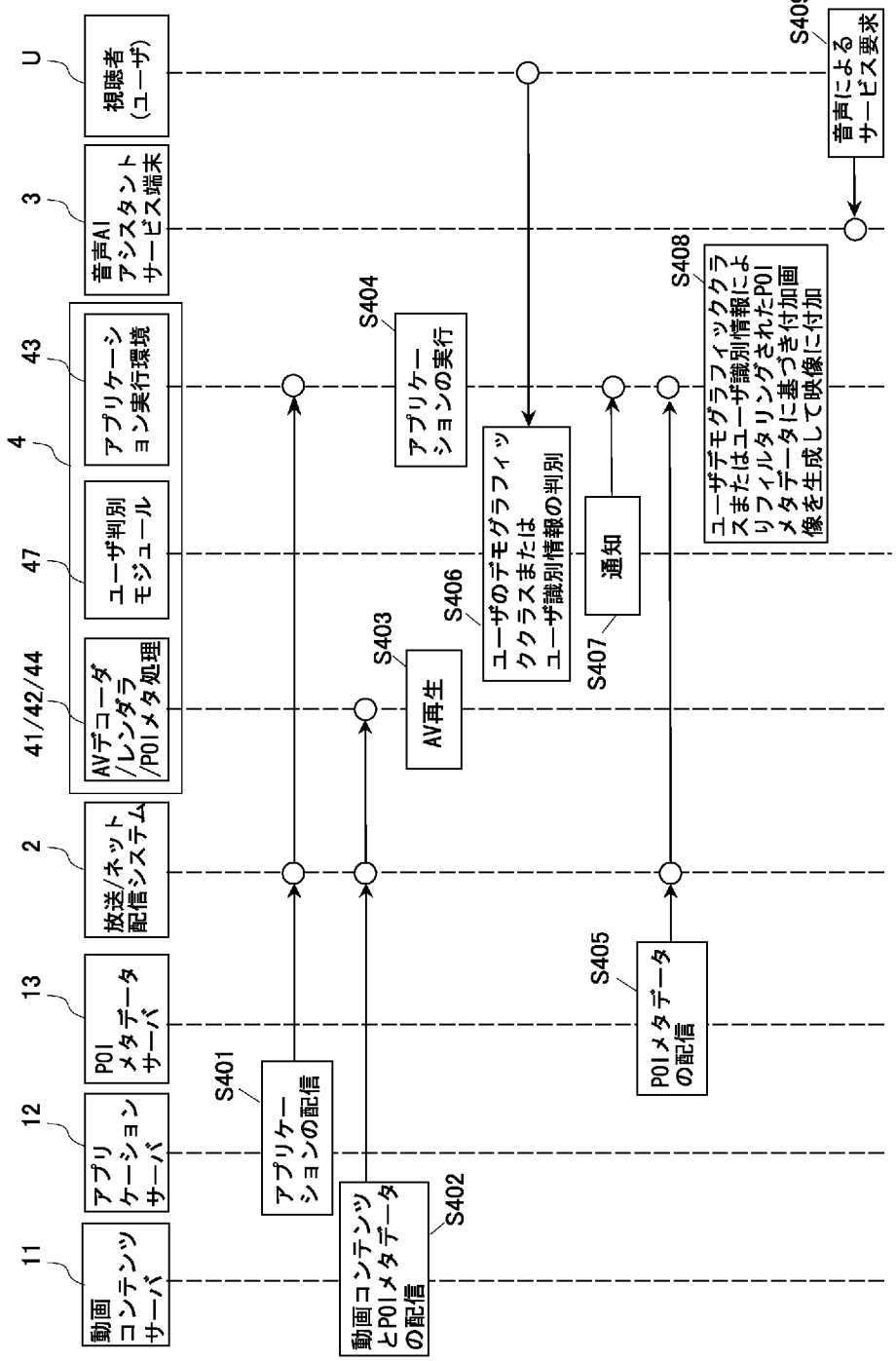
[図7]



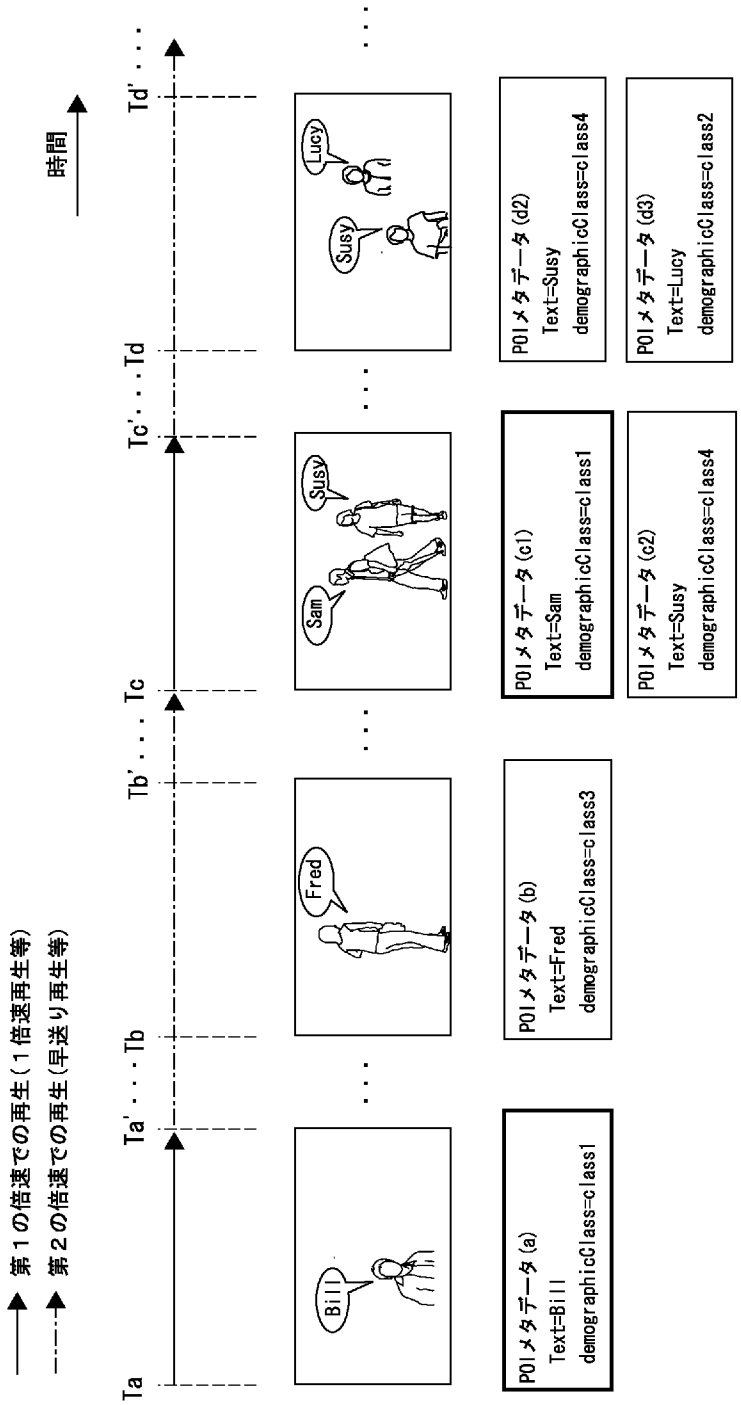
[図8]


51
Bill
Fred
52
53
Lucy
54
J3
J2
J1
55
Susy
Tom
J4
J5
付加画像の提示制限
51
Bill
J3
J2
J1
J4
J5

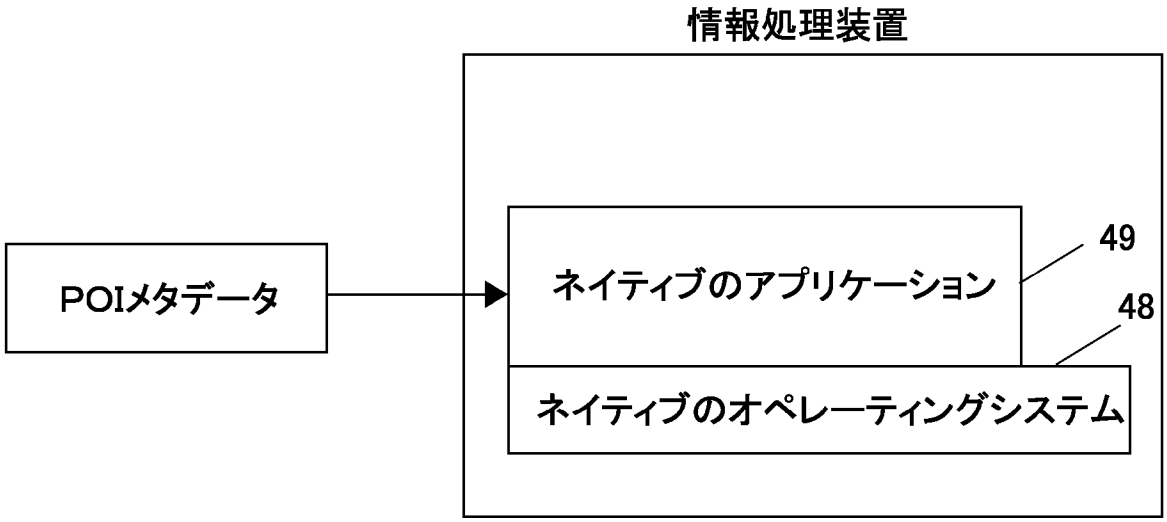
[図9]


U
視聴者
(ユーザ)
3
音声AI
アシスタント
サービス端末
4
43
アプリケーション実行環境
47
ユーザ判別
モジュール
41/42/44
AVデコーダ
/レンダラ
/POIメタ処理
2
放送/ネット
配信システム
13
POI
メタデータ
サーバ
12
アプリ
ケーション
サーバ
11
動画
コンテンツ
サーバ
S401
アプリケーション
の配信
S402
動画コンテンツ
とPOIメタデータ
の配信
S403
AV再生
S404
アプリケーション
の実行
S405
POIメタデータ
の配信
S406
ユーザのデモグラフィック
クラスまたは
ユーザ識別情報の判別
S407
通知
S408
ユーザデモグラフィック
クラスまたはユーザ識別情報によ
りフィルタリングされたPOI
メタデータに基づき付加画
像を生成して映像に付加
S409
音声による
サービス要求

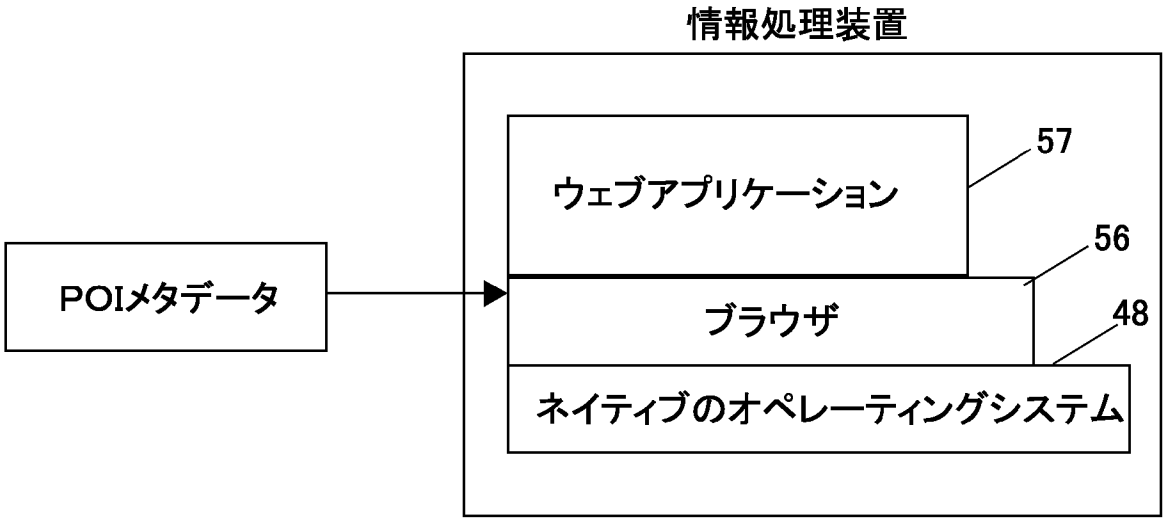
[図10]


第１の倍速での再生（１倍速再生等）
第２の倍速での再生（早送り再生等）
時間
Ta
Ta'・・・Tb
Tb'・・・・Tc
Tc'・・・Td
Td'・・・・
Bill
Fred
Sam
Susy
Susy
Lucy
POIメタデータ(a)
Text=Bill
demographicClass=class1
POIメタデータ(b)
Text=Fred
demographicClass=class3
POIメタデータ(c1)
Text=Sam
demographicClass=class1
POIメタデータ(c2)
Text=Susy
demographicClass=class4
POIメタデータ(d2)
Text=Susy
demographicClass=class4
POIメタデータ(d3)
Text=Lucy
demographicClass=class2

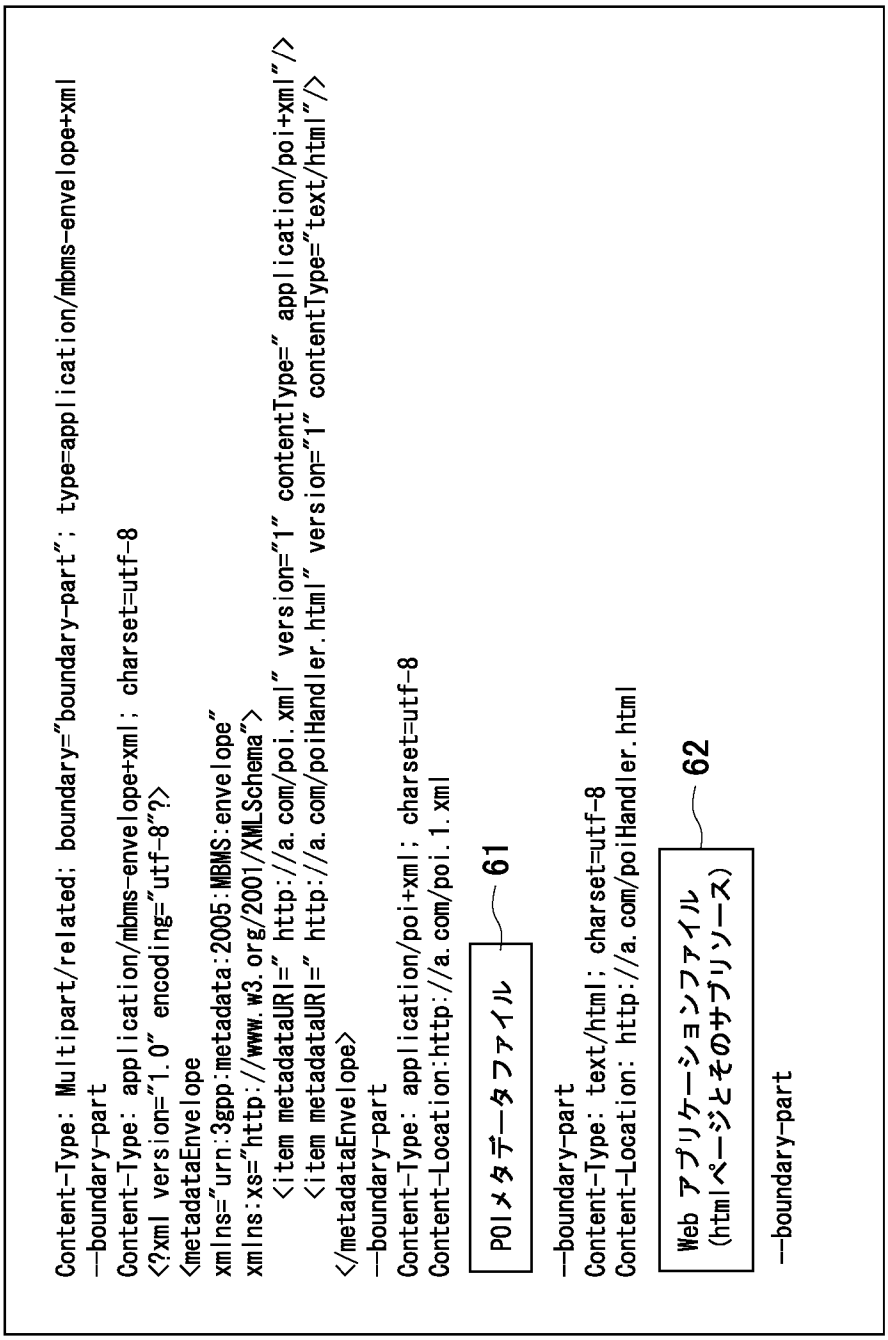
[図11]


情報処理装置
POIメタデータ
ネイティブのアプリケーション
49
48
ネイティブのオペレーティングシステム

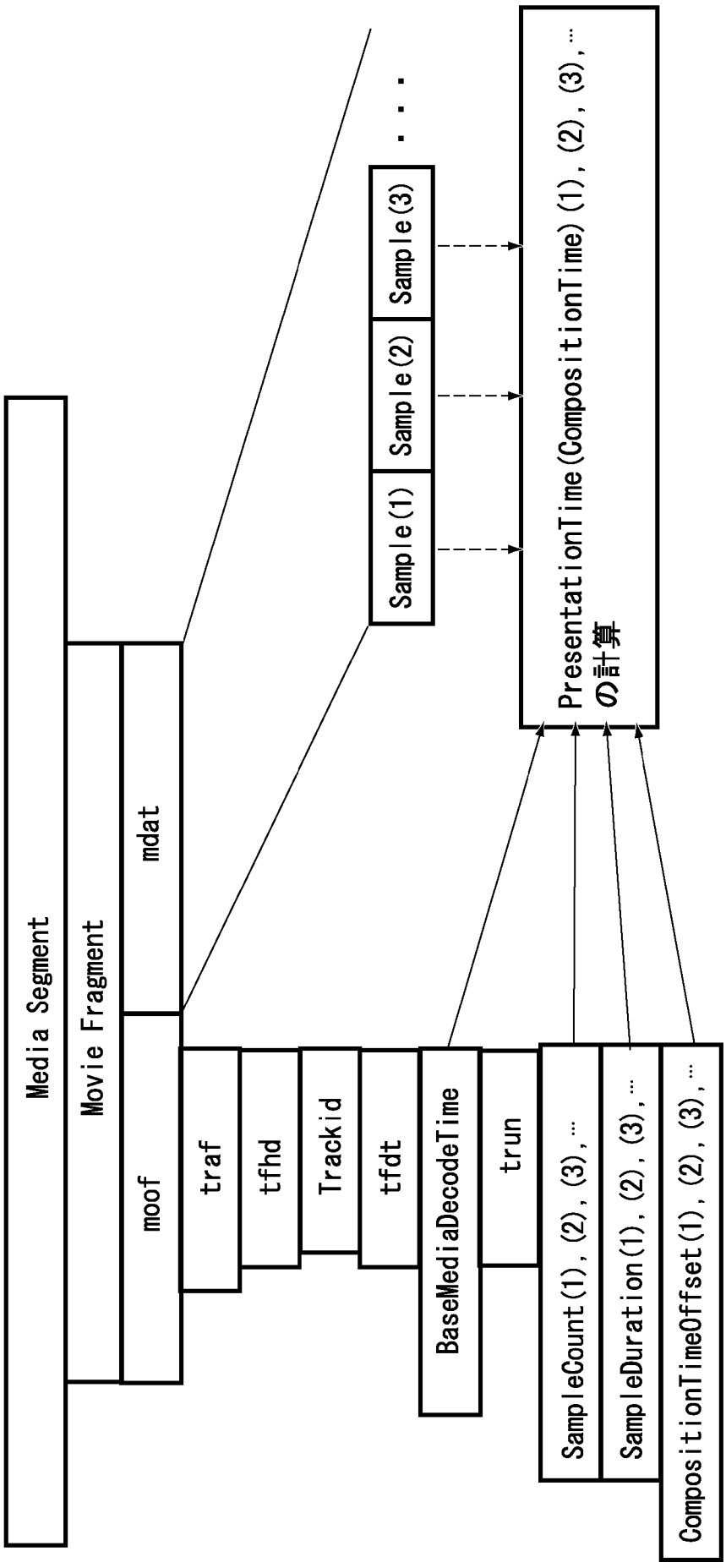
[図12]


情報処理装置
POIメタデータ
ウェブアプリケーション
57
56
ブラウザ
48
ネイティブのオペレーティングシステム

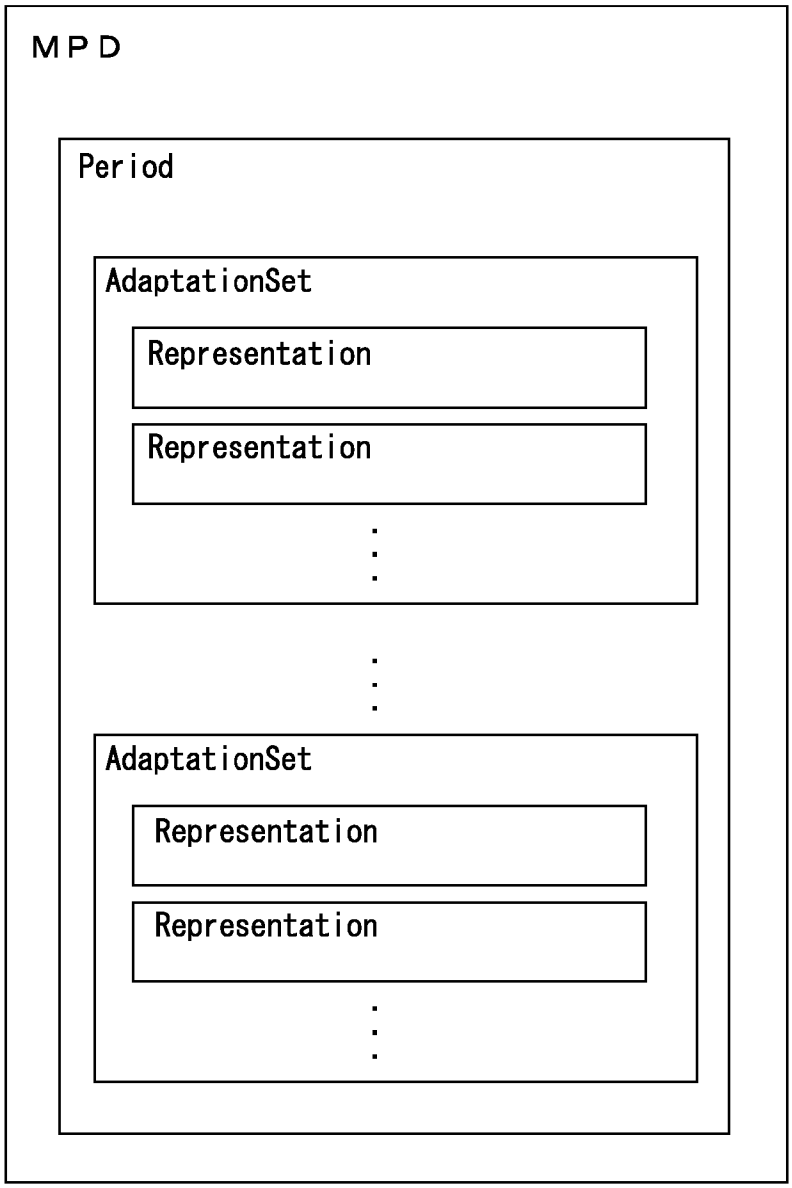
[図13]


Content-Type: Multipart/related; boundary="boundary-part"; type=application/mbms-envelope+xml
--boundary-part
Content-Type: application/mbms-envelope+xml; charset=utf-8
<?xml version="1.0" encoding="utf-8"?>
<metadataEnvelope
xmlns="urn:3gpp:metadata:2005:MBMS:envelope"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
<item metadataURI=" http://a.com/poi.xml" version="1" contentType=" application/poi+xml"/>
<item metadataURI=" http://a.com/poiHandler.html" version="1" contentType="text/html"/>
</metadataEnvelope>
--boundary-part
Content-Type: application/poi+xml; charset=utf-8
Content-Location:http://a.com/poi.1.xml
POIメタデータファイル
61
--boundary-part
Content-Type: text/html; charset=utf-8
Content-Location: http://a.com/poiHandler.html
Web アプリケーションファイル
(htmlページとそのサブリソース)
62
--boundary-part

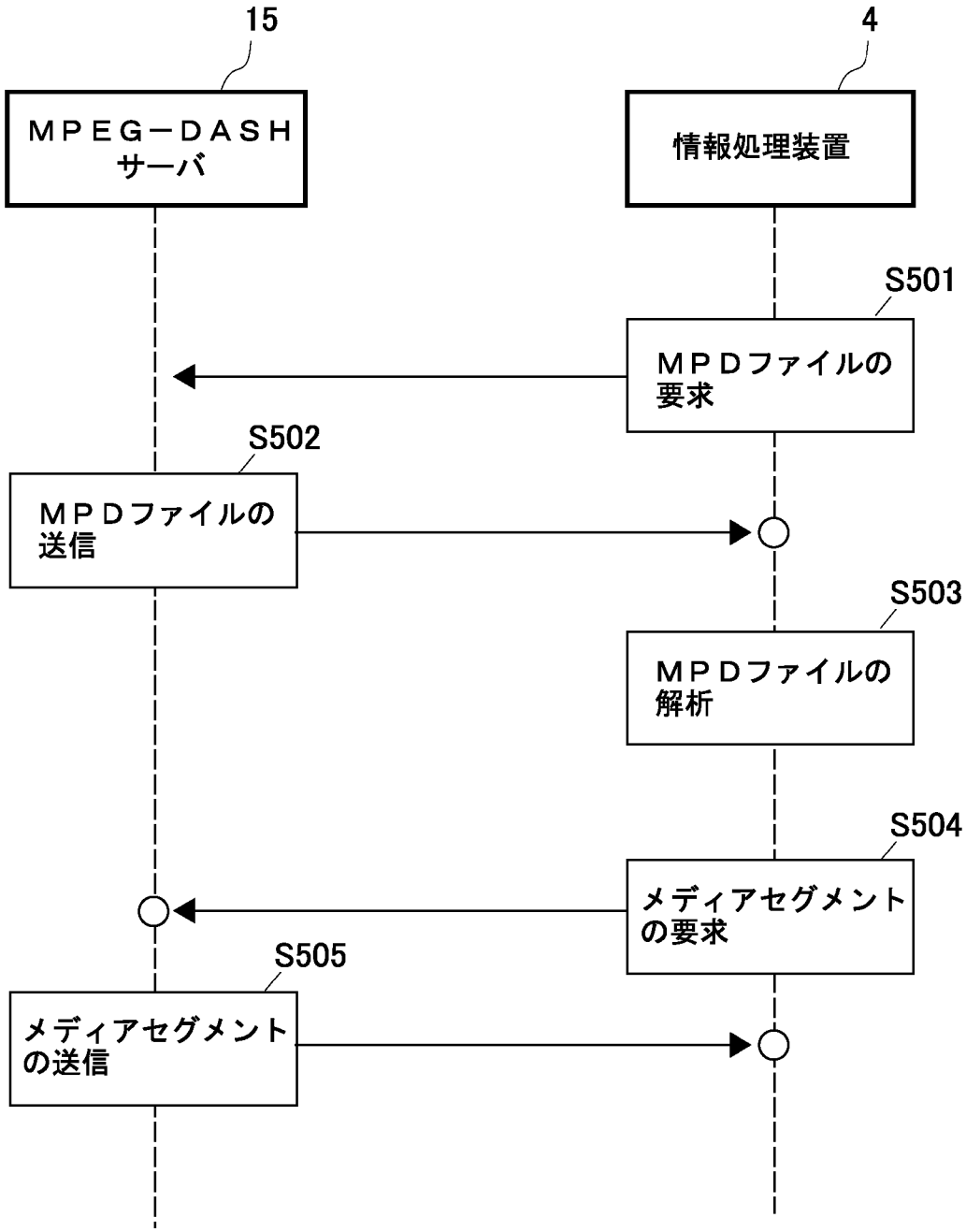
[図14]


Media Segment
Movie Fragment
moof
mdat
traf
tfhd
Trackid
tfdt
BaseMediaDecodeTime
trun
SampleCount(1),(2),(3),…
SampleDuration(1),(2),(3),…
CompositionTimeOffset(1),(2),(3),…
Sample(1)
Sample(2)
Sample(3)
…
PresentationTime(CompositionTime)(1),(2),(3),…の計算

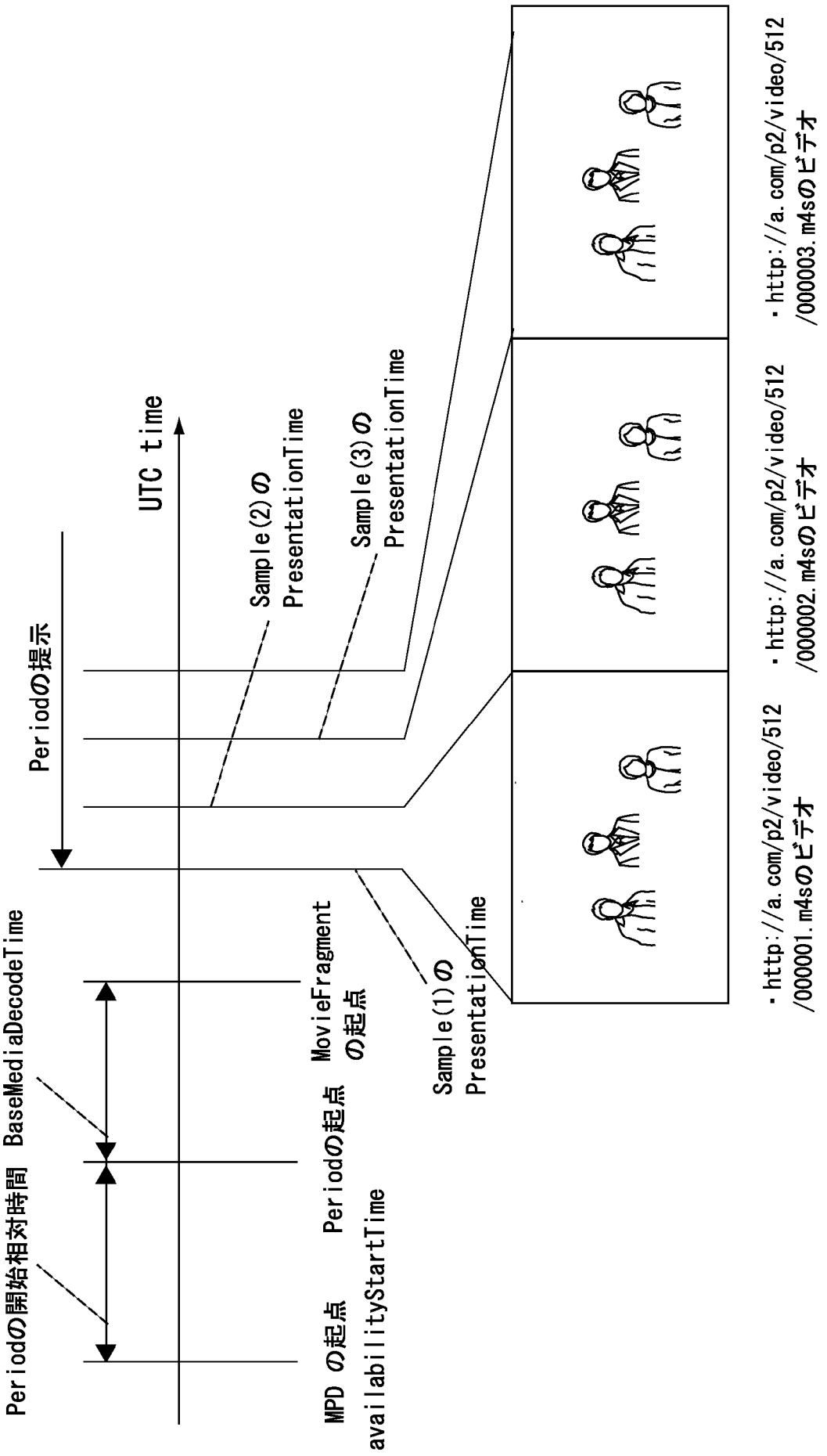
[図15]


MPD
Period
AdaptationSet
Representation
Representation
AdaptationSet
Representation
Representation

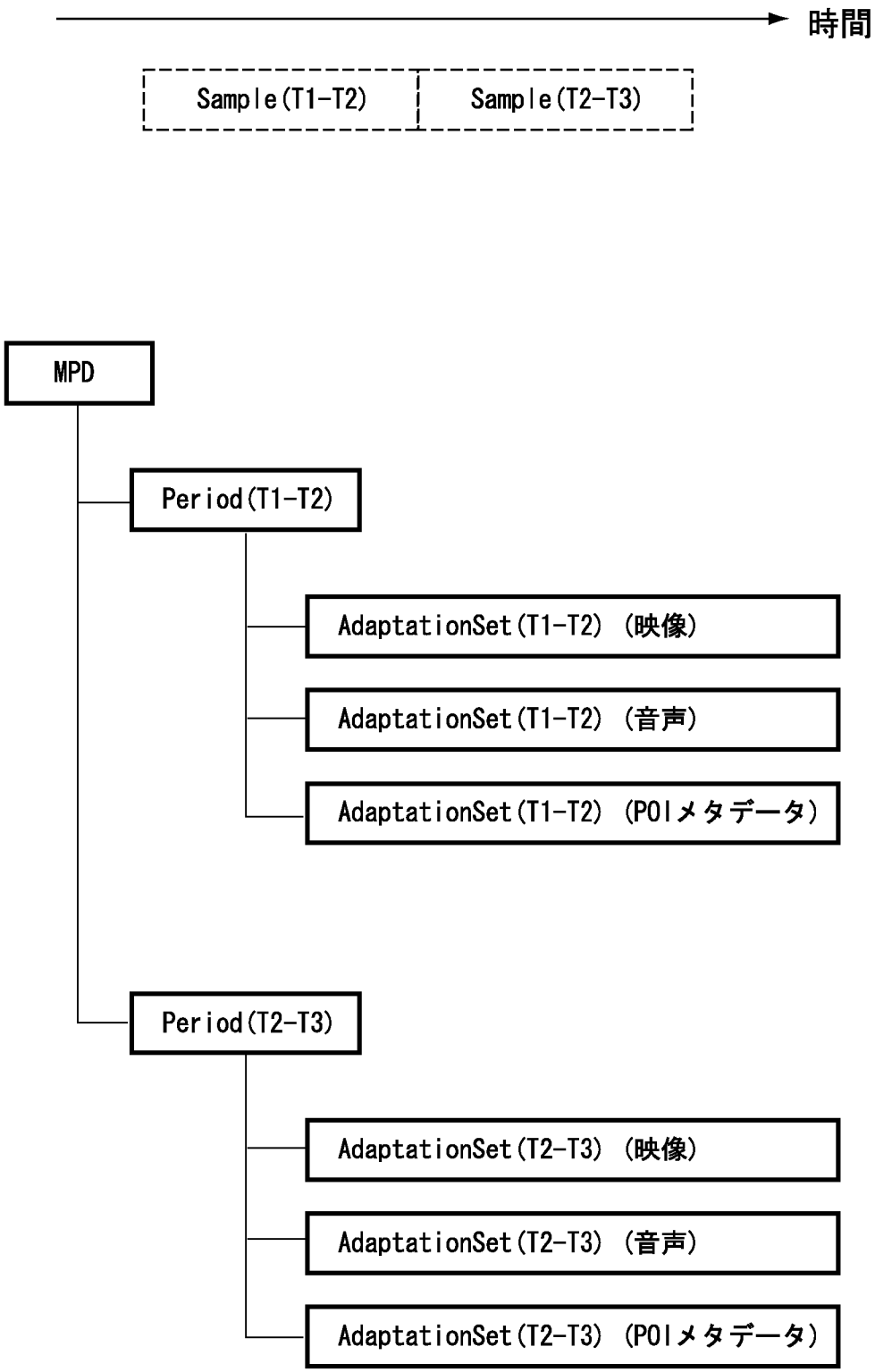
[図16]


15
4
ＭＰＥＧ－ＤＡＳＨ
サーバ
情報処理装置
S501
ＭＰＤファイルの
要求
S502
ＭＰＤファイルの
送信
S503
ＭＰＤファイルの
解析
S504
メディアセグメント
の要求
S505
メディアセグメント
の送信

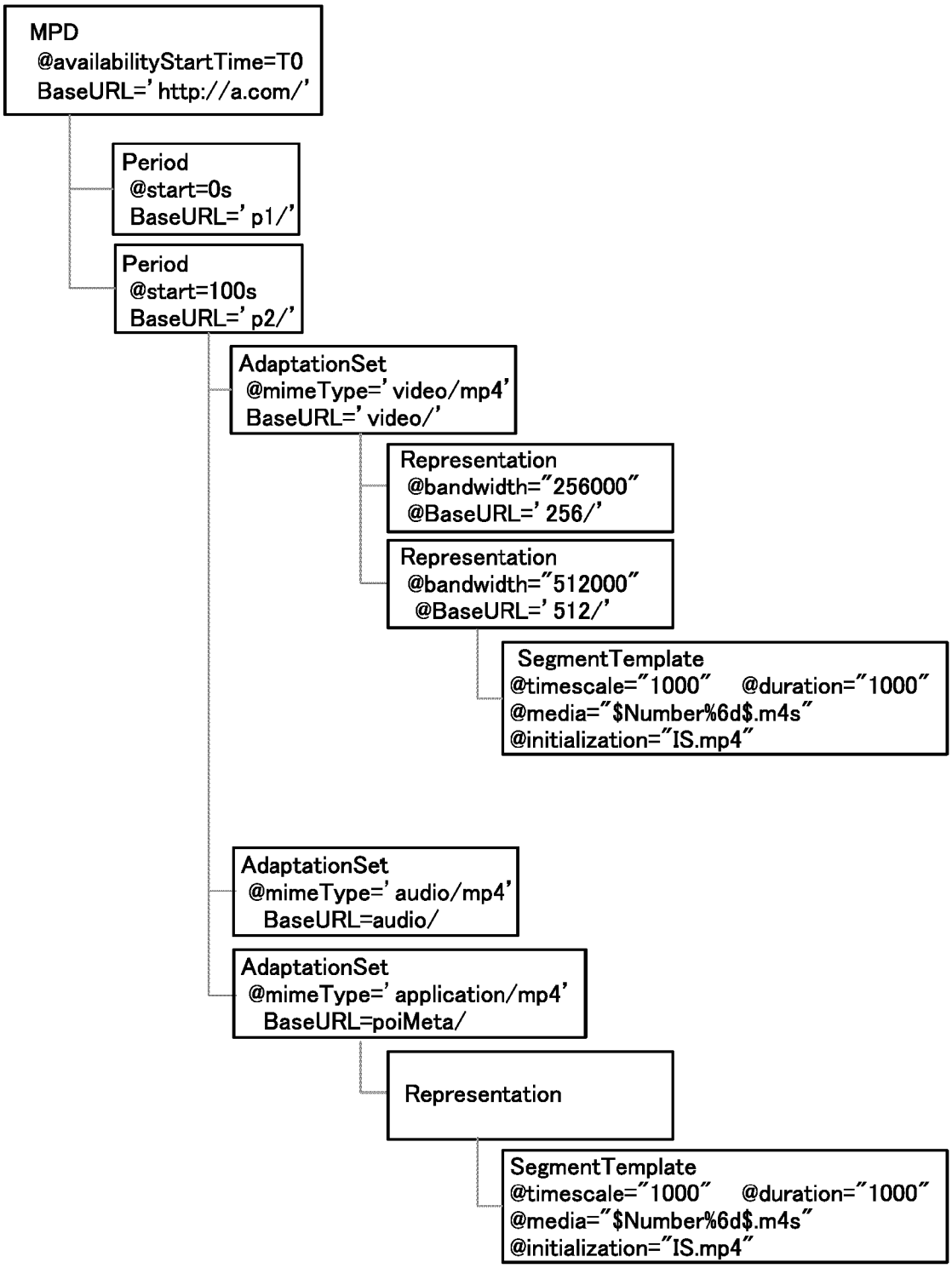
[図17]


UTC time
Periodの提示
Sample(2)の PresentationTime
Sample(3)の PresentationTime
Sample(1)の PresentationTime
Periodの開始相対時間 BaseMediaDecodeTime
MPD の起点 Periodの起点 MovieFragment の起点
availabilityStartTime
・http://a.com/p2/video/512 /000001.m4sのビデオ
・http://a.com/p2/video/512 /000002.m4sのビデオ
・http://a.com/p2/video/512 /000003.m4sのビデオ

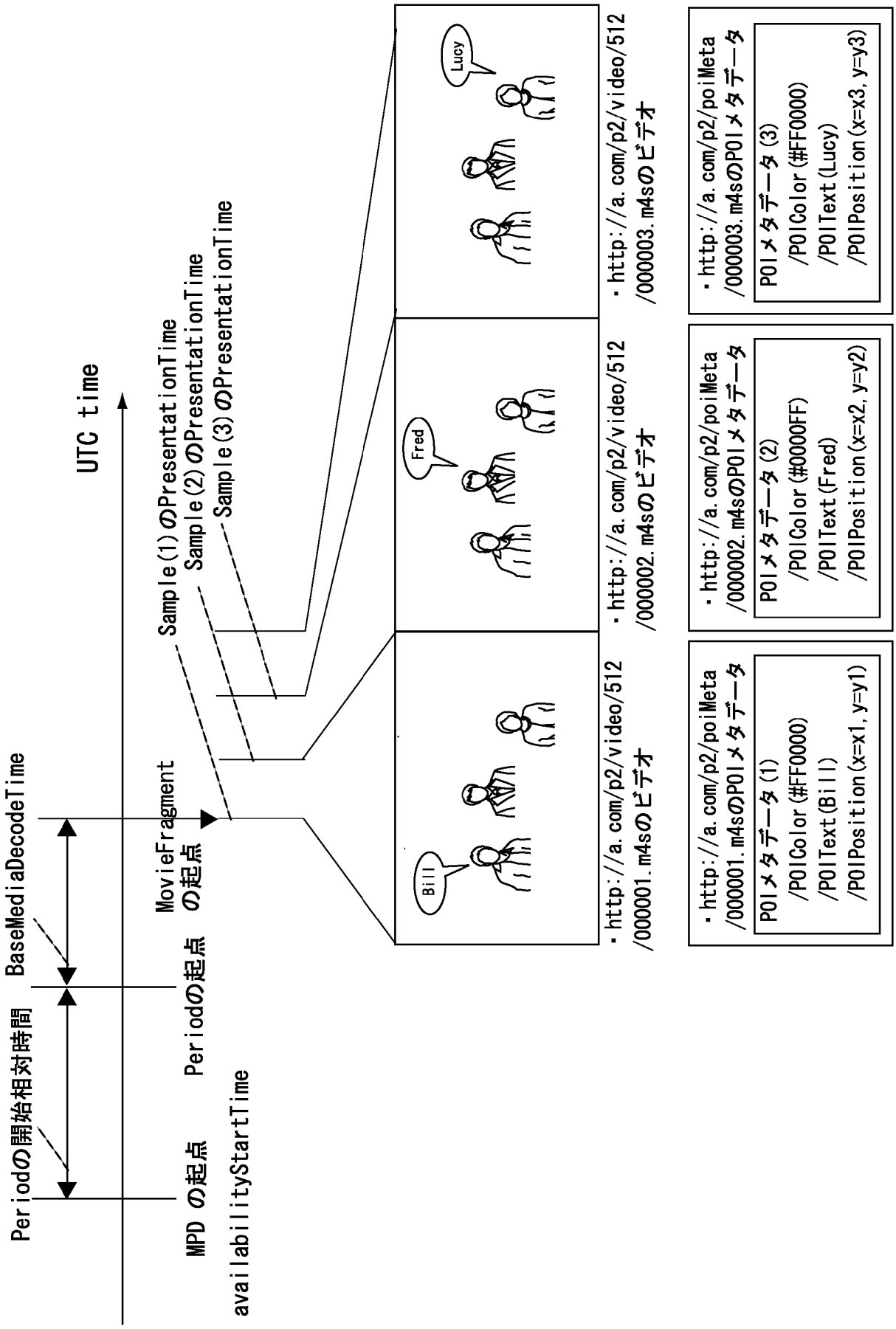
[図18]


時間
Sample(T1-T2)
Sample(T2-T3)
MPD
Period(T1-T2)
AdaptationSet(T1-T2) (映像)
AdaptationSet(T1-T2) (音声)
AdaptationSet(T1-T2) (POIメタデータ)
Period(T2-T3)
AdaptationSet(T2-T3) (映像)
AdaptationSet(T2-T3) (音声)
AdaptationSet(T2-T3) (POIメタデータ)

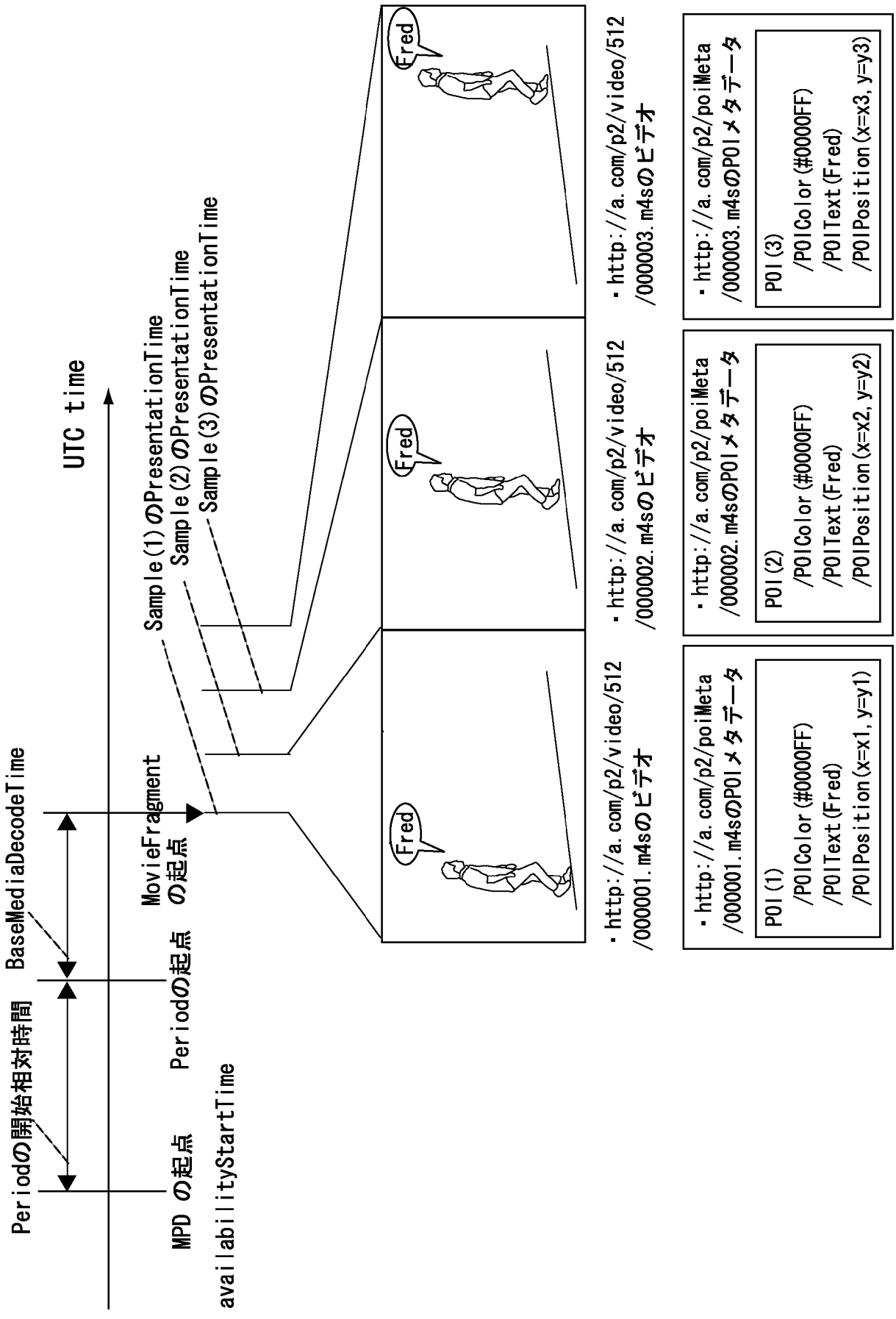
[図19]


MPD
@availabilityStartTime=T0
BaseURL=' http://a.com/'
Period
@start=0s
BaseURL=' p1/'
Period
@start=100s
BaseURL=' p2/'
AdaptationSet
@mimeType=' video/mp4'
BaseURL=' video/'
Representation
@bandwidth="256000"
@BaseURL=' 256/'
Representation
@bandwidth="512000"
@BaseURL=' 512/'
SegmentTemplate
@timescale="1000" @duration="1000"
@media="$Number%6d$.m4s"
@initialization="IS.mp4"
AdaptationSet
@mimeType=' audio/mp4'
BaseURL=audio/
AdaptationSet
@mimeType=' application/mp4'
BaseURL=poiMeta/
Representation
SegmentTemplate
@timescale="1000" @duration="1000"
@media="$Number%6d$.m4s"
@initialization="IS.mp4"

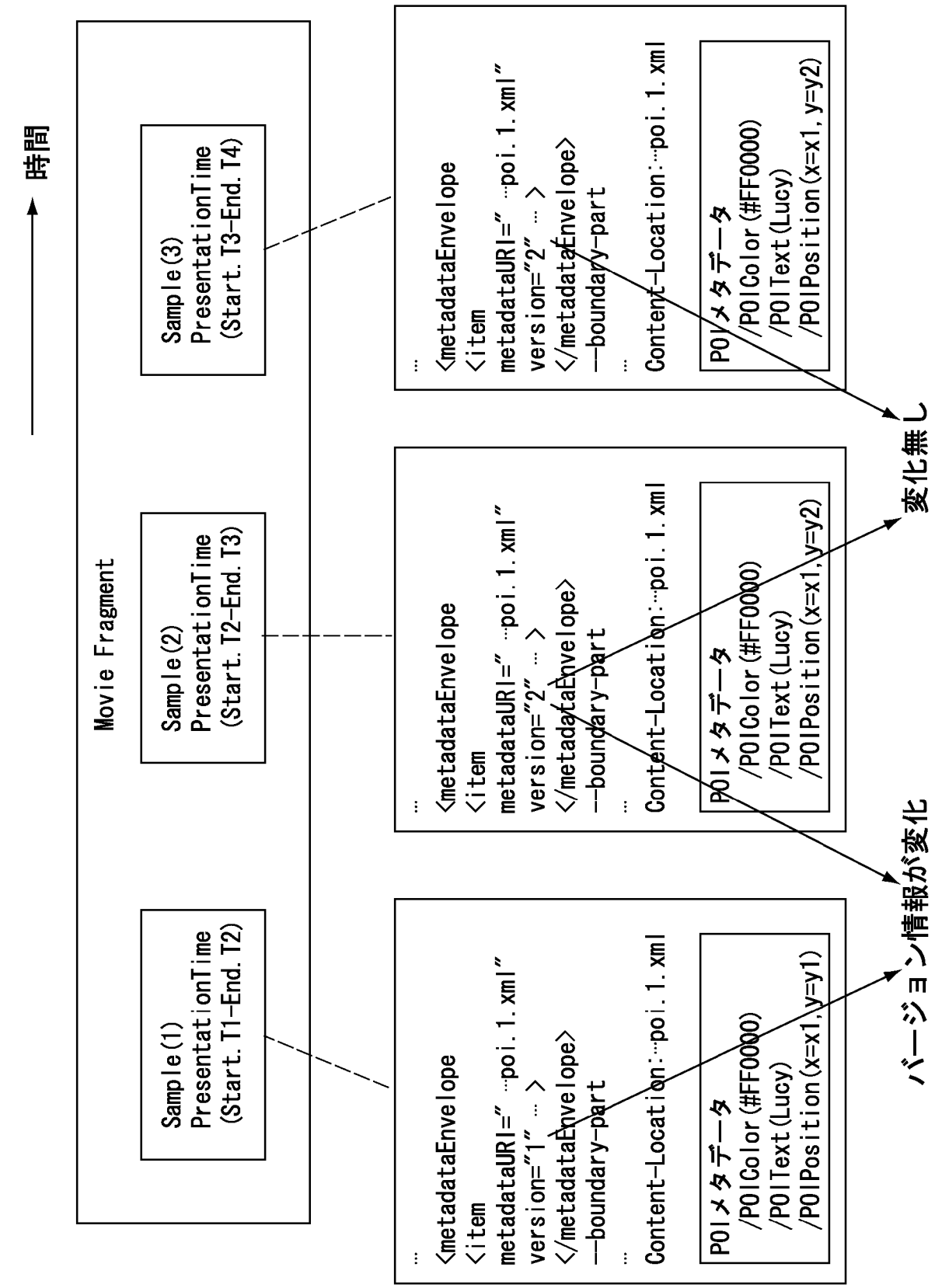
[図20]


UTC time
Periodの開始相対時間
BaseMediaDecodeTime
MPD の起点
Periodの起点
availabilityStartTime
MovieFragment の起点
Sample(1)のPresentationTime
Sample(2)のPresentationTime
Sample(3)のPresentationTime
Bill
Fred
Lucy
・http://a.com/p2/video/512/000001.m4sのビデオ
・http://a.com/p2/video/512/000002.m4sのビデオ
・http://a.com/p2/video/512/000003.m4sのビデオ
・http://a.com/p2/poiMeta/000001.m4sのPOIメタデータ
POIメタデータ(1)
/POIColor(#FF0000)
/POIText(Bill)
/POIPosition(x=x1,y=y1)
・http://a.com/p2/poiMeta/000002.m4sのPOIメタデータ
POIメタデータ(2)
/POIColor(#0000FF)
/POIText(Fred)
/POIPosition(x=x2,y=y2)
・http://a.com/p2/poiMeta/000003.m4sのPOIメタデータ
POIメタデータ(3)
/POIColor(#FF0000)
/POIText(Lucy)
/POIPosition(x=x3,y=y3)

[図21]


UTC time
Periodの開始相対時間
BaseMediaDecodeTime
MPD の起点
Periodの起点
availabilityStartTime
MovieFragment の起点
Sample(1)のPresentationTime
Sample(2)のPresentationTime
Sample(3)のPresentationTime
Fred
・http://a.com/p2/video/512/000001.m4sのビデオ
・http://a.com/p2/poiMeta/000001.m4sのPOIメタデータ
POI(1)
/POIColor(#0000FF)
/POIText(Fred)
/POIPosition(x=x1,y=y1)
・http://a.com/p2/video/512/000002.m4sのビデオ
・http://a.com/p2/poiMeta/000002.m4sのPOIメタデータ
POI(2)
/POIColor(#0000FF)
/POIText(Fred)
/POIPosition(x=x2,y=y2)
・http://a.com/p2/video/512/000003.m4sのビデオ
・http://a.com/p2/poiMeta/000003.m4sのPOIメタデータ
POI(3)
/POIColor(#0000FF)
/POIText(Fred)
/POIPosition(x=x3,y=y3)

[図22]


時間
Movie Fragment
Sample(1)
PresentationTime
(Start. T1-End. T2)
Sample(2)
PresentationTime
(Start. T2-End. T3)
Sample(3)
PresentationTime
(Start. T3-End. T4)
…
<metadataEnvelope
<item
metadataURI=" …poi. 1. xml"
version="1" … >
</metadataEnvelope>
—boundary-part
…
Content-Location:…poi. 1. xml
POIメタデータ
/POIColor(#FF0000)
/POIText(Lucy)
/POIPosition(x=x1, y=y1)
…
<metadataEnvelope
<item
metadataURI=" …poi. 1. xml"
version="2" … >
</metadataEnvelope>
—boundary-part
…
Content-Location:…poi. 1. xml
POIメタデータ
/POIColor(#FF0000)
/POIText(Lucy)
/POIPosition(x=x1, y=y2)
…
<metadataEnvelope
<item
metadataURI=" …poi. 1. xml"
version="2" … >
</metadataEnvelope>
—boundary-part
…
Content-Location:…poi. 1. xml
POIメタデータ
/POIColor(#FF0000)
/POIText(Lucy)
/POIPosition(x=x1, y=y2)
バージョン情報が変化
変化無し

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2019/008140

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl. H04N21/435(2011.01)i, G06F3/0481(2013.01)i, G06F3/16(2006.01)i, G10L15/00(2013.01)i, G10L15/22(2006.01)i, H04N21/488(2011.01)i, H04N21/84(2011.01)i
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
Int.Cl. H04N21/435, G06F3/0481, G06F3/16, G10L15/00, G10L15/22, H04N21/488, H04N21/84

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	1971-2019
Registered utility model specifications of Japan	1996-2019
Published registered utility model applications of Japan	1994-2019

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2002-369180 A (KYOCERA CORP.) 20 December 2002, paragraphs [0023]-[0024], [0029]-[0030] (Family: none)	1-3, 11-13
Y		4-7
A		8-10
Y	JP 2015-154195 A (NEC PERSONAL COMPUTERS, LTD.) 24 August 2015, paragraph [0019] (Family: none)	4-7

☒ 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 13 May 2019 (13.05.2019)	Date of mailing of the international search report 28 May 2019 (28.05.2019)
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2019/008140

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2012-249156 A (SONY CORP.) 13 December 2012, paragraph [0033] & EP 2530675 A2, paragraph [0033] & US 2012/0308202 A1	5-7

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１９／００８１４０

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

Int.Cl.　H04N21/435(2011.01)i, G06F3/0481(2013.01)i, G06F3/16(2006.01)i, G10L15/00(2013.01)i, G10L15/22(2006.01)i, H04N21/488(2011.01)i, H04N21/84(2011.01)i

Ｂ．　調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

Int.Cl.　H04N21/435, G06F3/0481, G06F3/16, G10L15/00, G10L15/22, H04N21/488, H04N21/84

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	１９７１－２０１９年
日本国実用新案登録公報	１９９６－２０１９年
日本国登録実用新案公報	１９９４－２０１９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 2002-369180 A（京セラ株式会社）2002.12.20, 段落 0023-0024, 0029-0030	1-3, 11-13
Y	（ファミリーなし）	4-7
A		8-10
Y	JP 2015-154195 A（ＮＥＣパーソナルコンピュータ株式会社）2015.08.24, 段落 0019 （ファミリーなし）	4-7

☑　Ｃ欄の続きにも文献が列挙されている。　　☐　パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「Ａ」特に関連のある文献ではなく、一般的技術水準を示すもの
「Ｅ」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「Ｌ」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「Ｏ」口頭による開示、使用、展示等に言及する文献
「Ｐ」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「Ｔ」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「Ｘ」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Ｙ」特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「＆」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
１３．０５．２０１９	２８．０５．２０１９

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
東京都千代田区霞が関三丁目４番３号

特許庁審査官（権限のある職員）　５Ｃ　２９５２
冨田　高史
電話番号　０３－３５８１－１１０１　内線　３５４１

C（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
Y	JP 2012-249156 A（ソニー株式会社）2012.12.13, 段落 0033 & EP 2530675 A2, 段落 0033 & US 2012/0308202 A1	5-7