



(51) International Patent Classification:
H04N 21/258 (2011.01)

(21) International Application Number:
PCT/IN2012/000800

(22) International Filing Date:
7 December 2012 (07.12.2012)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **HEWLETT-PACKARD DEVELOPMENT COMPANY, L. P** [US/US]; 11445 Compaq Center Drive West, Houston, Texas, Texas 77070 (US).

(72) Inventors; and

(71) Applicants : **MADHVANATH, Sriganesh** [IN/IN]; 24, Salarpuria Arena, Adugodi, Hosur Road, Bangalore 560030 (IN). **VENNELAKANTI, Ramadevi** [IN/IN]; 24, Salarpuria Arena, Adugodi, Hosur Road, Bangalore 560030 (IN). **DEY, Prasenjit** [IN/IN]; 24, Salarpuria Arena, Adugodi, Hosur Road, Bangalore 560030 (IN).

(74) Agents: **LAKSHMIKUMARAN, Varadachari** et al.; B-6/10, Safdarjung Enclave, New Delhi 110029 (IN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): 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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (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, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

(54) Title: CREATING MULTIMODAL OBJECTS OF USER RESPONSES TO MEDIA

300

CAPTURING A MULTIMODAL USER RESPONSE TO THE MEDIA OBJECT 02

MAPPING THE MULTIMODAL USER RESPONSE TO A FILE OF THE MEDIA OBJECT 04

CREATING A MULTIMODAL OBJECT INCLUDING THE MAPPED MULTIMODAL USER RESPONSE AND THE MEDIA OBJECT 306

Fig. 3

(57) Abstract: Creating a multimodal object of a user response to a media object can include capturing a multimodal user response to the media object, mapping the multimodal user response to a file of the media object, and creating a multimodal object including the mapped multimodal user response and the media object.

WO 2014/087415 A1

CREATING MULTIMODAL OBJECTS OF USER RESPONSES TO MEDIA

Background

[0001] People can view media, such as photographs, video, and television content on a variety of devices, both individually and in social settings. Responses to the viewed media can be multimodal in nature. For instance, responses to media can include facial gestures, hand gestures, speech, and non-speech sounds.

Brief Description of the Drawings

[0002] Figure 1 is a flow chart illustrating an example of a process for creating a multimodal object of a user response to a media object according to the present disclosure.

[0003] Figure 2 illustrates an example of a multimodal object according to the present disclosure.

[0004] Figure 3 is a block diagram illustrating an example of a method for creating a multimodal object of a user response to a media object according to the present disclosure.

[0005] Figure 4 illustrates an example of a system including a computing device according to the present disclosure.

Detailed Description

[0006] Consumer responses to media, such as a media object, can be useful for a variety of purposes. For instance, captured responses can be shared with others (e.g., friends and family) who are remotely located, can be used to identify what advertisers to associate with a particular media object, and/or can be used to determine an effectiveness of a media object (e.g., positive reaction to an advertisement).

[0007] Media and/or media objects can be viewed and/or packaged in a number of ways. For instance, Internet media sites (e.g., YouTube and Flickr) and social network sites (e.g., Facebook, Twitter, and GooglePlus) allow users (e.g., consumers) to comment on media objects posted by others. The comments tend to be textual in nature and can be studied responses instead of spontaneous responses and/or interactions. Such textual responses, for example, tend to be limited in emotional content.

[0008] In some instances, a real-time camera-based audience measurement system can be used to understand how an online and/or road billboard advertisement is being received. Such systems can count how many people have viewed the billboard and potentially analyze the demographics of viewers.

[0009] Further, video screen capture, sometimes referred to as screencast, can contain audio narration based off of a digital recording of a computer screen output. Screencasts can be used to demonstrate and/or teach the use of software features, in education to integrate technology into curriculum, and for capturing seminars and/or presentations. Screencasts tend to capture purposeful screen activity and audio narration of the presenter, rather than spontaneous responses of the viewers.

[0010] However, internet media sites and social network sites, real-time camera-based audience measurement systems, and video screen captures tend to be limited as they cannot capture multiple aspects of the user response such as the tone of a response, a gesture of a user's face and/or head, and/or something that is pointed to in the media object. In contrast, examples in accordance with the

present disclosure can be used to capture and format a multimodal user response to a media object as the response occurs. The resulting multimodal user response can, for instance, be a real-time user response including multiple modalities of the response.

[0011] Examples of the present disclosure may include methods, systems, and computer-readable and executable instructions and/or logic. An example method for creating a multimodal object of a user response to a media object can include capturing a user response to the media object, mapping the user response to a file of the media object, and creating a multimodal object including the mapped user response and the media object.

[0012] In the following detailed description of the present disclosure, reference is made to the accompanying drawings that form a part hereof, and in which is shown by way of illustration how examples of the disclosure may be practiced. These examples are described in sufficient detail to enable those of ordinary skill in the art to practice the examples of this disclosure, and it is to be understood that other examples may be utilized and the process, electrical, and/or structural changes may be made without departing from the scope of the present disclosure.

[0013] The figures herein follow a numbering convention in which the first digit or digits correspond to the drawing figure number and the remaining digits identify an element or component in the drawing. Similar elements or components between different figures may be identified by the use of similar digits. Elements shown in the various examples herein can be added, exchanged, and/or eliminated so as to provide a number of additional examples of the present disclosure.

[0014] In addition, the proportion and the relative scale of the elements provided in the figures are intended to illustrate the examples of the present disclosure, and should not be taken in a limiting sense. As used herein, the designators "N" and "P" particularly with respect to reference numerals in the drawings, indicate that a number of the particular feature so designated can be included with a number of examples of the present disclosure. Also, as used

herein, "a number of" an element and/or feature can refer to one or more of such elements and/or features.

[0015] Figure 1 is a flow chart illustrating an example of a process 110 for creating a multimodal object of a user response to a media object according to the present disclosure. A media object, as used herein, can be a file of a video, audio (e.g., music and/or speech), photograph, slideshow of photographs, and/or a document, among many other files. A user can include a consumer, a viewing user, and/or an associated user (e.g., friend, family member, and/or co-worker of the creator of the media object), among many other people that may view a media object.

[0016] A user can view a media object on a computing device 130. For instance, the computing device 130 can include a browser 118 and/or a media application 114. The media application 114 can run on a computing device 130, for example. A browser 118 can include an application (e.g., computer-executable instructions) for retrieving, presenting, and traversing information resources (e.g., domains, images, video, and/or other content) on the Internet. The media application 114 can include a native and/or a non-native application. A native application can include an application (e.g., computer-readable instructions) that operates under the same operating system and/or operating language as the computing device.

[0017] A non-native application can include an application (e.g., computer-readable instructions) that is web-based (e.g., operating language is a browser-rendered language, such as Hyper Text Markup Language combined with JavaScript) and/or not developed for a particular operating system (e.g., Java application and/or a browser 118). A non-native application and/or native application 114 may use a plug-in 116, in some instances, to support creation and/or playback of a multimodal object 120 from media objects stored locally. A plug-in 116, as used herein, can include computer-executable instructions that enable customizing the functionality of an application (e.g., to play a media object,

access components of the computing device 130 to create a multimodal object 120, and/or playback the multimodal object 120).

[0018] The media object can, for instance, be stored locally on the computing device 130 of the user and/or can be stored externally. For instance, a media object can be stored externally in a cloud system, a social media network, and/or many other external sources and/or external systems. A media object stored on an external source and/or system can be accessed and/or viewed by the user using the browser 118 and/or the Internet, for example.

[0019] The process 110 can include capturing a user response to a media object. A user response, as used herein, can include a reaction and/or interaction of the user to viewing the media object.

[0020] A user response can include a multimodal user response. A modality, as used herein, can include a particular way for information to be presented and/or communicated to and/or by a human. A multimodal user response can include multiple modalities of responses by a user.

[0021] For instance, as illustrated in the example of Figure 1, multiple modalities of user responses can include sound 112-1, gestures 112-2, touch 112-3, user context 112-N, and/or other responses. Sound 112-1 can include words spoken, laughter, sighs, and/or other noises. Gestures 112-2 can include hand gestures, face gestures, head gestures, and/or other body gestures of the user. Touch 112-3 can include point movements (e.g., as discussed further in Figure 2), among other movements. User context 112-N can, for instance, include a level of attention, an identity of the user, and/or facial expression of the viewing user, among other context.

[0022] The multimodal user responses 112-1, 112-2, 112-3, ..., 112-N can be captured using a computing device 130 of the user. For instance, the multimodal user responses 112-1, ..., 112-N can be captured using a native and/or non-native application (e.g., media application 114 and/or browser 118), a plug-in 116, a camera, a microphone, a display, and/or other hardware and/or software

components (e.g., computer-executable instructions) of the user computing device 130. The captured user responses can include user response data.

[0023] The captured multimodal user responses can, in some examples of the present disclosure, be user configurable. For instance, a user can be provided a user-configurable selection of types of user response data to capture prior to capturing the multimodal user responses and/or viewing the media object. The user-configurable selection can be provided in a user interface. For instance, the user interface can include a display allowing a user to select the types of user responses to capture. The modalities of user responses captured can be in response to the user selection.

[0024] The captured multimodal user responses can be mapped to a file of the media object based on a common timeline. The common timeline, as used herein, can include the timeline of the media object. For example, mapping the multimodal user responses can include processing and/or converting the user responses into sub-portions, annotating the processed responses with reference to a time and/or place in the media object, and mapping each sub-portion of the user responses to the time and/or place in the media object (e.g., as discussed further in Figure 3).

[0025] Using the mapped user responses, a multimodal object 120 can be created. The multimodal object 120 can include the mapped user responses and the media object. For instance, the multimodal object 120 can be a multilayer multimodal object. A multilayer multimodal object can include each modality of the user's responses 112-1, ..., 112-N and the media object on a separate layer of the multilayer multimodal object.

[0026] In various examples of the present disclosure, the media object can be stored externally (e.g., in a cloud system). A media object stored externally can be used and/or viewed to create a multimodal object 122 using a browser 118 and a plug-in 116. A user can grant the plug-in 116 permission to access components of the user computing system 130 to capture user response data and/or create a multimodal object 122. The multimodal object 122 created using a media object

stored externally can include a link that can be shared, for example. For instance, the link can be embedded as a part of the multimodal object 122 and/or include an intrinsic attribute of the multimodal object 122.

[0027] In some examples of the present disclosure, a multimodal object 122 created using a media object stored externally can include a set of user response data. The set of user response data can include an aggregation of multiple users responses to the media object stored externally. The multimodal object 122 can accumulate and/or aggregate the multiple users' responses with the media object over time.

[0028] In various examples of the present disclosure, the set of user response data and/or a user response to the media object can include multiple co-present users' responses to the media object. Multiple co-present users can include multiple users viewing and/or interacting over media in a co-present manner. Co-present, as used herein, can include synchronously (e.g., viewing and/or interacting at a common time). In some examples, synchronously can include simultaneously. The multiple co-present users' responses to a media object can be shared, for example. For instance, the multiple co-present users' responses can be shared with an end-user and/or stored externally in an external system.

[0029] For instance, multiple co-present users can include a co-located group of users (e.g., multiple users located in the same location) and/or non co-located group of users (e.g., viewing at the same time using the Internet). Multiple users that are co-located can include a group of users located around a system sharing the media object. For instance, user response data captured from the multiple co-located users can be stored on an external system (e.g., a cloud system) and/or internal system (e.g., a device associated with the multiple users).

[0030] A non co-located group of users can view a media object on the Internet (e.g., a whiteboard application) while each user in the group is located at different points and/or locations. User response data from the multiple non co-located group of users can be aggregated automatically using an external system

(e.g., aggregate in a cloud system as captured) and/or locally on each of the user's computing systems using the external system (e.g., synchronize each user's computing system and aggregate in the external system).

[0031] In some examples of the present disclosure, each response of a user, among a non co-located group of users, to a media object can be captured non-synchronously (e.g., asynchronously), and can be processed to and/or into a synchronous multimodal object. As an example, user A can be located at location I, user B can be located at location II, and user C can be located at location III. User A, user B, and user C can view the media object at their respective locations at separate and/or different times. Each user's response (e.g., user A, user B, and user C) can be captured at a computing device associated with the respective user and mapped to a file of the media object based on a common timeline (e.g., timeline of the media object). A multimodal object can be created on and/or use an external system (e.g., cloud system) by aggregating each mapped user response to the file of the media object to create a multiuser multimodal object including each user's mapped multimodal user response and the media object.

[0032] The multimodal object created (e.g., multimodal object in a cloud 122 and/or multimodal object 120 internally stored) can be distributed to an end-user. Distribution can include sharing, sending, and/or otherwise providing the multimodal object to an end-user to be viewed. An end-user, as used herein, can include a creator of the media object (e.g., company, organization, and/or third-party to a company and/or organization), a company and/or organization, a system (e.g., cloud system, social network, Internet, etc.), and/or many other persons that may benefit from viewing the multimodal object.

[0033] In various examples of the present disclosure, a multimodal object 122 created, stored, and/or accessed from an external system can track and/or aggregate responses to the media object and/or the multimodal media object from an external system user. An external system user can include a social network user, a cloud system user, and/or Internet user, among many other system users. The external system user can include a user on the external system the multimedia

object 122 is created, stored, and/or accessed from a separate and/or different external system.

[0034] A multimodal object 122 stored on an external system (e.g., cloud system) can be accessed and/or viewed (e.g., played) by a number of end-users. For instance, a number of end-users that are located in a number of locations can view the multimodal object 122 on a number of devices. Each device among the number of devices can be associated with an end-user among the number of end-users. Further, if the media object is stored on the external system (e.g., a photograph shared on a photograph sharing site), it may be easier to capture multiple users' responses to create a multimodal object 122 than if the media object were stored on an internal system because the media object can be accessed by the number of end-users.

[0035] For instance, a multimodal object 122 created from a media object stored externally can include captured social network responses to the media object. The social network responses can be captured and incorporated into the media object. Social network responses and/or external system responses can include comments on the media object and can be treated as audio comments from a user, for example. In some examples, if the external system user has granted permission to access the external system user's computing device (e.g., webcam, microphone, etc.), a full multimodal response can be captured. If the external system user has not granted permission, text comments can be captured.

[0036] The distributed multimodal object 120,122 can be viewed by the end-user. The end-user can view the multimodal object 120,122 using a native and/or non-native media application 124, a plug-in 126, and/or a browser 128 on a computing device 132 of and/or associated with the end-user. Viewing the multimodal object 120, 122 can include a synchronous view of each layer of the multimodal object (e.g., the media object and each modality of the user response) based on a common timeline.

[0037] Figure 2 illustrates an example of a multimodal object 234 according to the present disclosure. A multimodal object 234, as illustrated by Figure 2, can

include captured user response data. The captured user response data can include multiple layers. For instance, each layer 236-1, 236-2, ..., 236-P, 238 can include one modality of a user response 236-1, ..., 236-P and/or the file of the media object 238 based on a common timeline 240.

[0038] The multimodal object 234 can be viewed by an end-user on a user interface (e.g., a display). For instance, the multimodal object 234 can be viewed, displayed, and/or played back to the end-user in a synchronous view of each layer 236-1, ..., 236-P, 238 of the multimodal object 234 to recreate the live interaction experience and/or response of the user.

[0039] A synchronous view can include display and/or play back of user response data captured (e.g., 236-1, ..., 236-P) and/or processed with the media object (e.g., 238) playing at the same time. For instance, the media object 238 can be rendered in a separate window. Mouse and/or other forms of point movements can be superimposed as pointers on the media object itself 238 to represent where the user has pointed. Point movements, as used herein, can include user movements and/or pointing toward a display (e.g., screen, touch screen, and/or mobile device screen) while a media object is playing. The point movements can be accomplished by moving a mouse, touching a display, and/or pointing from a distance (e.g., sensed using a depth camera). The point movements can be in reference to a media object (e.g., a point of interest in the media object). The point movements captured can be represented in the created multimodal object as a separate layer 236-2 with the point movements represented by reference to a space on the media object pointed to.

[0040] In some examples, the user response data can be processed and/or converted to a text format and the text can be displayed. For instance, audio and/or other input modalities captured can be processed, converted, and/or displayed as subtitles and/or text at the bottom of the screen (e.g., as illustrated by the text "bored", "amazed", and "happy" of layer 236-1). The text can be displayed with added animation (e.g., virtual characters as illustrated in 236-1) and/or

converted into other forms (e.g., synthesized laughter to represent laughing as illustrated in 236-P).

[0041] The user response data, in various examples, can be processed, converted, and/or displayed in sub-portions. For instance, the sub-portions can be represented as text and/or can include the actual sub-portions of the interaction data collected. The sub-portions, in some examples, can be processed in separate layers. The layers of modality 236-1, ..., 236-P can each include video, audio, and/or screenshots of the user response (e.g., live pictures and/or video of the user responding to the video and/or live audio recordings), among other representations.

[0042] Figure 3 is a block diagram illustrating an example of a method 300 for creating a multimodal object of a user response to a media object according to the present disclosure. At 302, the method 300 can include capturing a multimodal user response to the media object. The multimodal user response can be recorded using a camera, microphone, and/or other hardware and/or software (e.g., executable instruction) components of a computing device of and/or associated with the user. The captured multimodal user response can include user response data, for instance.

[0043] A multimodal user response to a media object can include multiple modalities of response. For example, response to media objects can include modalities such as facial gestures, hand gestures, speech sounds, and/or non-speech sounds.

[0044] At 304, the method 300 can include mapping the multimodal user response to a file of the media object. Mapping can, for instance, be based on a common timeline. For example, mapping can include annotating each multimodal user response to a media object with a reference to the media object. For instance, a user response to a media object can be annotated with reference to a particular time (e.g., point in time) in the media object that each response occurred and/or reference to a place in the media object (e.g., a photograph in a slideshow).

[0045] In some examples of the present disclosure, the captured multimodal user response data can be processed. For instance, the captured user response data can be converted to multiple sub-portions, to labels, and/or text. The multiple sub-portions can, for example, be used to remove silences (e.g., empty space in the user response data) in the user response to reduce storage space as compared to the complete user response data. The labels and/or text can be obtained and/or converted from the user response data using speech-to-text convertors, facial detection and facial expression recognition, and/or hand gesture interpreters, for instance. For instance, a face can be identified from a set of registered faces. The registered faces can include faces corresponding to frequent viewers (e.g., family and friends).

[0046] The converted sub-portions, labels, and/or text can be derived from the complete user response data and can be annotated with timestamps and/or references to a specific and/or particular place (e.g., photograph, time, and/or image) corresponding to when the sub-portion occurred with respect to the media object viewed.

[0047] As an example, a media object can include a photographic slideshow of two pictures. A user response to a first picture can be converted and/or processed to a first sub-portion (e.g., cut into a piece and/or snippet) and can be annotated with a reference to the first photograph. The user response to a second picture can be converted and/or processed to a second sub-portion and can be annotated with a reference to the second photograph. If the user does not have a response during viewing of the media object for a period of time (e.g., between the first photograph and the second photograph), the user response data containing no response can be removed from the captured user response data. Using the annotated references, the multimodal user response to the first picture can be mapped to the first picture and the multimodal user response to the second picture can be mapped to the second picture.

[0048] At 306, the method 300 can include creating a multimodal object including the mapped multimodal user response and the media object. The

multimodal object can include a multilayer file of each modality of the user response data associated with the file of the media object. For instance, a multilayer file of each modality can include a file containing multiple channels of the user response data that can be layered and based on a common timeline (e.g., the timeline of the media object).

[0049] Figure 4 illustrates an example of a system including a computing device 442 according to the present disclosure. The computing device 442 can utilize software, hardware, firmware, and/or logic to perform a number of functions.

[0050] The computing device 442 can be a combination of hardware and program instructions configured to perform a number of functions. The hardware, for example can include one or more processing resources 444, computer-readable medium (CRM) 448, etc. The program instructions (e.g., computer-readable instructions (CRI)) can include instructions stored on the CRM 448 and executable by the processing resources 444 to implement a desired function (e.g., capturing a user response to the media object, etc.).

[0051] CRM 448 can be in communication with a number of processing resources of more or fewer than 444. The processing resources 444 can be in communication with a tangible non-transitory CRM 448 storing a set of CRI executable by one or more of the processing resources 444, as described herein. The CRI can also be stored in remote memory managed by a server and represent an installation package that can be downloaded, installed, and executed. The computing device 442 can include memory resources 446, and the processing resources 444 can be coupled to the memory resources 446.

[0052] Processing resources 444 can execute CRI that can be stored on an internal or external non-transitory CRM 448. The processing resources 444 can execute CRI to perform various functions, including the functions described in Figures 1-3.

[0053] The CRI can include a number of modules 450, 452, 454, and 456. The number of modules 450, 452, 454, and 456 can include CRI that when executed by the processing resources 444 can perform a number of functions.

[0054] The number of modules 450, 452, 454, and 456 can be sub-modules of other modules. For example, the multimodal map module 452 and the creation module 454 can be sub-modules and/or contained within a single module. Furthermore, the number of modules 450, 452, 454, and 456 can comprise individual modules separate and distinct from one another.

[0055] A capture module 450 can comprise CRI and can be executed by the processing resources 444 to capture a multimodal user response to the media object. In some examples of the present disclosure, the multimodal user response can be captured using an application. The application can, for instance, include a native application, non-native application, and/or a plug-in. The multimodal user response can be captured using a camera, microphone, and/or other hardware and/or software components of a computing device of and/or associated with the user. The native application and/or plug-in can request use of the camera and/or microphone, for example.

[0056] A multimodal map module 452 can comprise CRI and can be executed by the processing resources 444 to convert the multimodal user response into a number of layered sub-portions, annotate each layered sub-portion with a reference to the media object, and map each layered sub-portion of the multimodal user response to a file of the media object based on a common timeline and the annotation to the media object. A layer can, for instance, include a modality of the multimodal user response and/or the file of the media object, for example.

[0057] A creation module 454 can comprise CRI and can be executed by the processing resources 444 to create a multimodal object including the mapped layered user response and the media object. In some examples, the creation module 454 can include instructions to aggregate multiple users' responses to the media object. The multiple users can be co-present. For instance, the multiple users' responses can be synchronous (e.g., users are co-located and/or viewing the media object in a synchronized manner) and/or asynchronous (e.g., users are non co-located, viewing the media object at different times, and/or the aggregation can occur using an external system).

[0058] A distribution module 456 can comprise CRI and can be executed by the processing resources 444 to send the multimodal object to an end-user. For instance, the end-user can include a company and/or organization, a third party to the company and/or organization, a viewing user (e.g., family and/or friend of the user), and/or a system (e.g., a cloud system, a social network, and a social media site). The distribution module 456 can, in some examples, include instructions to store and/or upload the multimodal object to an external system (e.g., cloud system and/or social network). In such examples, the media object may be stored on the external system, in addition to the multimodal object.

[0059] In some examples, a system for creating a multimodal object of a user response to a media object can include a display module. A display module can comprise CRI and can be executed by the processing resources 444 to display the multimodal object using a native application and/or a plug-in of the computing device of and/or associated with the end-user. The multimodal object can be sent, for instance, to the end-user. The end-user can playback and/or view a received multimodal object. The playback and/or view can include a synchronous view and/or display of each layer of the multimodal object based on the common timeline. Each layer can include a modality of the user interaction data which can be displayed as text, sub-titles, animation, real audio and/or video, synthesized audio, among many other formats.

[0060] A non-transitory CRM 448, as used herein, can include volatile and/or non-volatile memory. Volatile memory can include memory that depends upon power to store information, such as various types of dynamic random access memory (DRAM), among others. Non-volatile memory can include memory that does not depend upon power to store information. Examples of non-volatile memory can include solid state media such as flash memory, electrically erasable programmable read-only memory (EEPROM), phase change random access memory (PCRAM), magnetic memory, and/or a solid state drive (SSD), etc., as well as other types of computer-readable media.

[0061] The non-transitory CRM 448 can be integral, or communicatively coupled, to a computing device, in a wired and/or a wireless manner. For example, the non-transitory CRM 448 can be an internal memory, a portable memory, a portable disk, or a memory associated with another computing resource (e.g., enabling CRIs to be transferred and/or executed across a network such as the Internet).

[0062] The CRM 448 can be in communication with the processing resources 444 via a communication path. The communication path can be local or remote to a machine (e.g., a computer) associated with the processing resources 444. Examples of a local communication path can include an electronic bus internal to a machine (e.g., a computer) where the CRM 448 is one of volatile, non-volatile, fixed, and/or removable storage medium in communication with the processing resources 444 via the electronic bus.

[0063] The communication path can be such that the CRM 448 is remote from the processing resources, (e.g., processing resources 444) such as in a network connection between the CRM 448 and the processing resources (e.g., processing resources 444). That is, the communication path can be a network connection. Examples of such a network connection can include a local area network (LAN), wide area network (WAN), personal area network (PAN), and the Internet, among others. In such examples, the CRM 448 can be associated with a first computing device and the processing resources 444 can be associated with a second computing device (e.g., a Java[®] server). For example, a processing resource 444 can be in communication with a CRM 448, wherein the CRM 448 includes a set of instructions and wherein the processing resource 444 is designed to carry out the set of instructions.

[0064] As used herein, "logic" is an alternative or additional processing resource to perform a particular action and/or function, etc., described herein, which includes hardware (e.g., various forms of transistor logic, application specific integrated circuits (ASICs), etc.), as opposed to computer executable instructions (e.g., software, firmware, etc.) stored in memory and executable by a processor.

[0065] The specification examples provide a description of the applications and use of the system and method of the present disclosure. Since many examples can be made without departing from the spirit and scope of the system and method of the present disclosure, this specification sets forth some of the many possible example configurations and implementations.

What is claimed:

1. A method for creating a multimodal object of a user response to a media object, comprising:
 - capturing a multimodal user response to the media object;
 - mapping the multimodal user response to a file of the media object; and
 - creating the multimodal object including the mapped multimodal user response and the media object.
2. The method of claim 1, including converting the multimodal user response to the media object to multiple sub-portions of the multimodal user response.
3. The method of claim 1, wherein capturing the multimodal user response to the media object includes using a browser on a computing device.
4. The method of claim 1, wherein capturing the multimodal user response to the media object includes using a media application on a computing device.
5. The method of claim 1, wherein capturing the multimodal user response includes capturing multiple co-present users' responses to the media object.
6. The method of claim 5, wherein capturing the multiple co-present users' responses to the media object includes aggregating each user response among the multiple co-present users' responses to the media object using an external system.
7. The method of claim 1, including requesting access to a component of a computing device of the user to capture the multimodal user response.

8. A non-transitory computer-readable medium storing a set of instructions executable by a processing resource, wherein the set of instructions can be executed by the processing resource to:

capture a multimodal user response to a media object;

annotate the captured multimodal user response with a reference to the media object;

map the multimodal user response to a file of the media object based on a common timeline and the annotation to the media object; and

create a multimodal object including the mapped multimodal user response and the media object.

9. The non-transitory computer-readable medium of claim 8, wherein the instructions executable by the processing resource include instructions to aggregate multiple multimodal user responses to the media object in the multimodal object.

10. The non-transitory computer-readable medium of claim 8, wherein the instructions executable by the processing resource include instructions to capture a response of an external system user to the media object and incorporate the response in the multimodal object.

11. The non-transitory computer-readable medium of claim 8, wherein the instructions executable by the processing resource include instructions to provide a user-configurable selection of types of user response data to capture.

12. A system for creating a multimodal object of a user response to a media object comprising:

a processing resource;

a memory resource coupled to the processing resource to implement:

a capture module including computer-readable instructions stored on the memory resource and executable by the processing resource to capture a multimodal user response to the media object;

a multimodal map module including computer-readable instructions stored on the memory resource and executable by the processing resource to:

convert the multimodal user response into a number of layered sub-portions;

annotate each layered sub-portion with a reference to the media object; and

map each layered sub-portion of the multimodal user response to a file of the media object based on a common timeline and the annotation to the media object;

a creation module including computer-readable instructions stored on the memory resource and executable by the processing resource to create the multimodal object including the mapped layered user response and the media object; and

a distribution module including computer-readable instructions stored on the memory resource and executable by the processing resource to send the multimodal object to an end-user.

13. The system of claim 12, wherein the distribution module includes instructions to upload the multimodal object to an external system.

14. The system of claim 12, wherein the system includes a display module including computer-readable instructions stored on the memory resource and executable by the processing resource to display the multimodal object using a native application.

15. The system of claim 12, the system includes a display module including computer-readable instructions stored on the memory resource and executable by

the processing resource to display the multimodal object to the end-user, wherein the display includes a synchronous view of each layer of the multimodal object based on the common timeline.

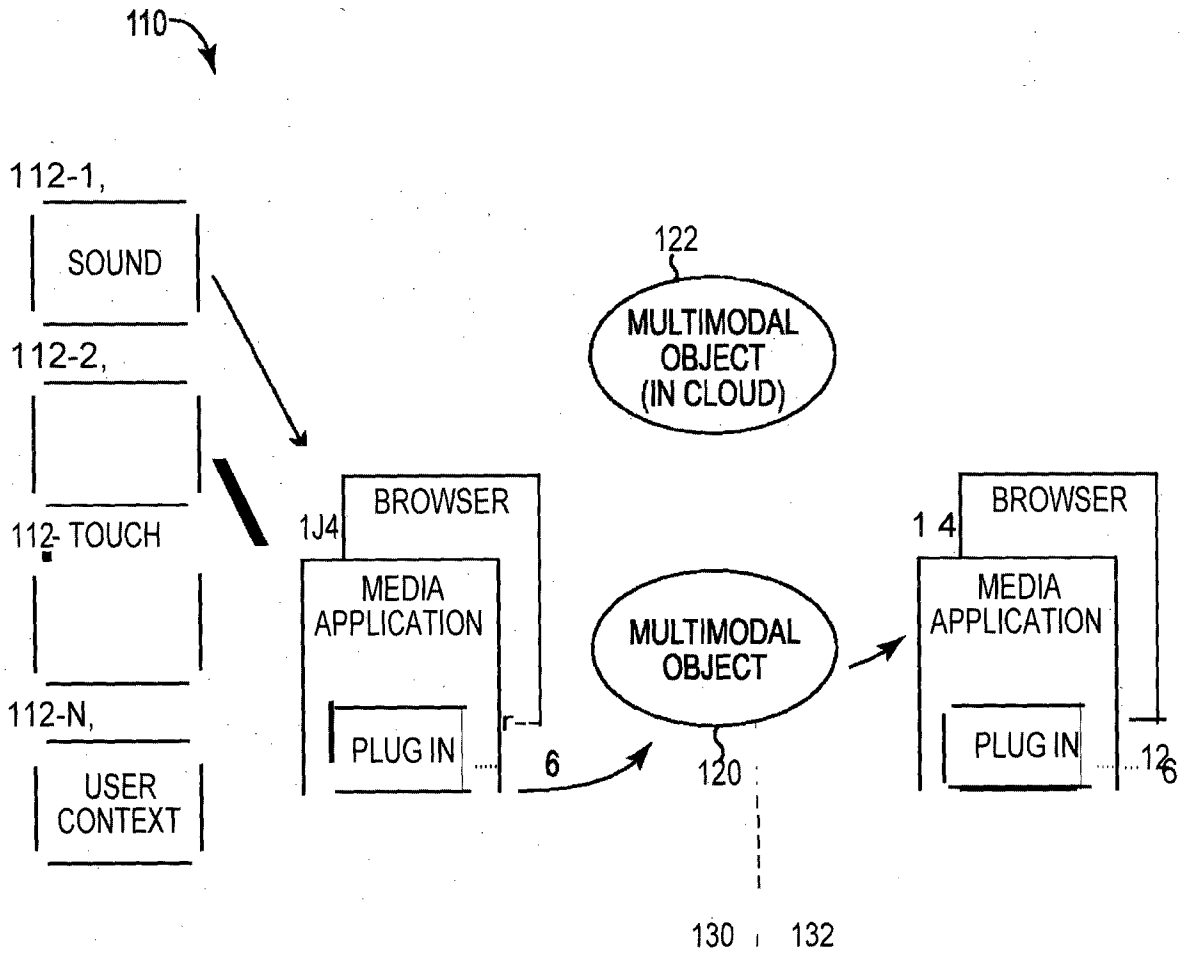
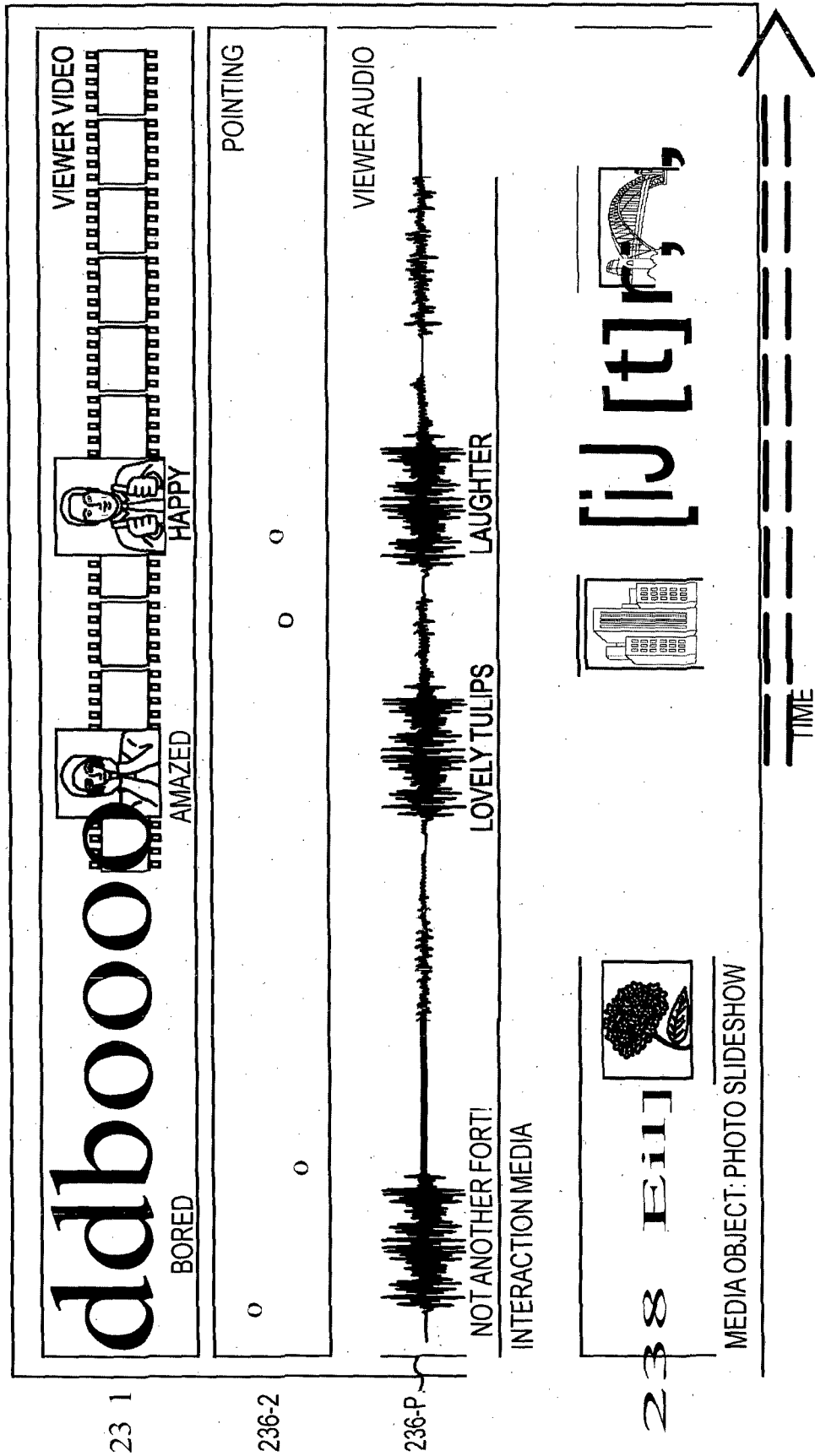


Fig. 1

234



2 Fig.2

300

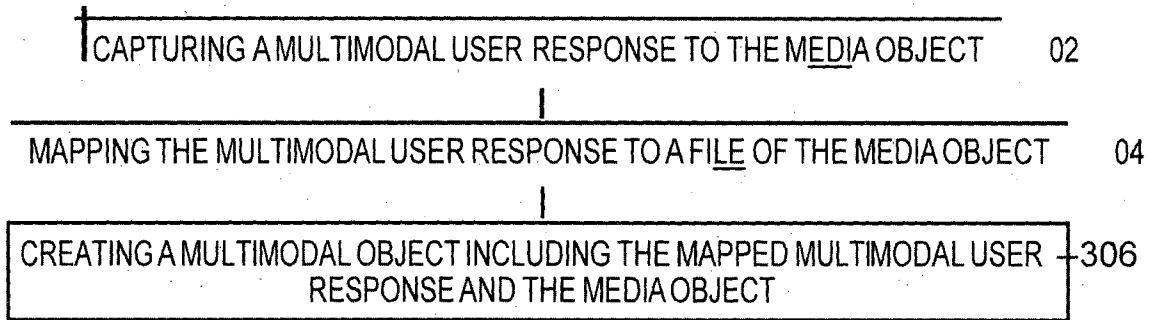


Fig. 3

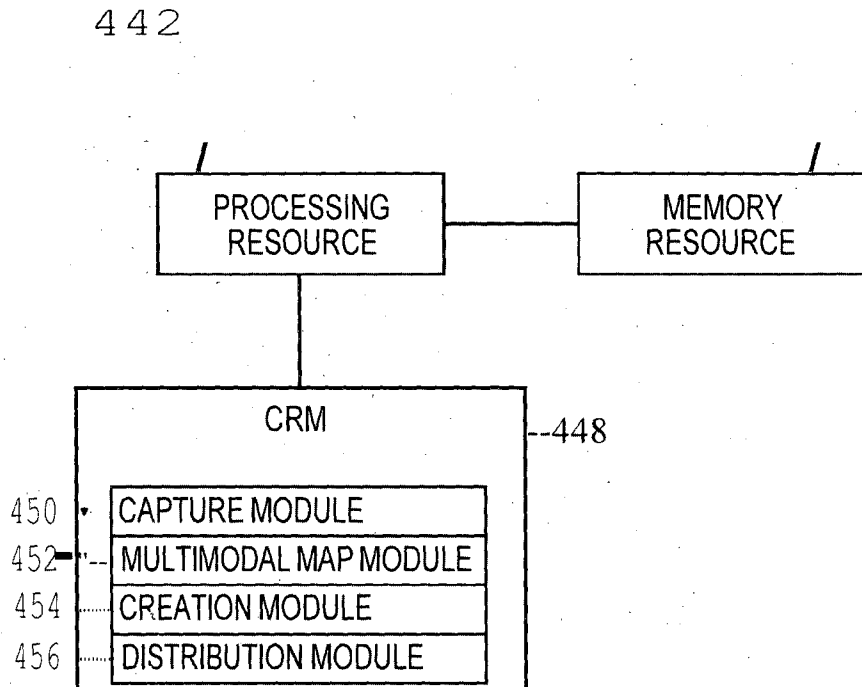


Fig. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IN2012/000800

A. CLASSIFICATION OF SUBJECT MATTER		
H04N 21/258 (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)		
IPC: H04L; G06F; H04N; G11B		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
CNPAT, WPI, EPODOC, IEEE, CNKI: media, multimedia, user, response, face, gesture, speech, voice, capture, track		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102522102 A (MICROSOFT CORPORATION) 27 June 2012 (27.06.2012) description, paragraphs [0003]-[0004] and [0016]-[0057]	1-15
X	CN 102572539 A (MICROSOFT CORPORATION) 11 July 2012 (11.07.2012) description, paragraphs [0003]-[0006], [0019]-[0060] and [0082]-[0114]	1-15
A	US 2010235739 A1 (APPLE INC.) 16 September 2010 (16.09.2010) the whole document	1-15
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents:	“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 “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 “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 “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) “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 “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 “&” document member of the same patent family	
Date of the actual completion of the international search 19 July 2013 (19.07.2013)	Date of mailing of the international search report 22 Aug. 2013 (22.08.2013)	
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer FU, Yuanyuan Telephone No. (86-10)62413558	

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/IN2012/000800

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 102522102 A	27.06.2012	US 2012093481 A1	19.04.2012
CN 102572539 A	11.07.2012	US 2012124604 A1	17.05.2012
US 2010235739 A1	16.09.2010	None	