MMEDIA 2021

The Thirteenth International Conferences on Advances in Multimedia

ISBN: 978-1-61208-839-6

April 18 - 22, 2021

MMEDIA 2021 Editors

Hiroshi Ishikawa, Tokyo Metropolitan University. Japan

MMEDIA 2021

Forward

The Thirteenth International Conference on Advances in Multimedia (MMEDIA 2021) aimed to provide an international forum by researchers, students, and professionals for presenting recent research results on advances in multimedia, mobile and ubiquitous multimedia and to bring together experts from both academia and industry for the exchange of ideas and discussion on future challenges in multimedia fundamentals, mobile and ubiquitous multimedia, multimedia ontology, multimedia user-centered perception, multimedia services and applications, and mobile multimedia.

The rapid growth of information on the Web, its ubiquity and pervasiveness makes the www the biggest repository. While the volume of information may be useful, it creates new challenges for information retrieval, identification, understanding, selection, etc. Investigating new forms of platforms, tools, principles offered by Semantic Web opens another door to enable humans programs, or agents to understand what records are about, and allows integration between domain-dependent and media-dependent knowledge. Multimedia information has always been part of the Semantic Web paradigm, but requires substantial effort to integrate both.

The new technological achievements in terms of speed and the quality of expanding and creating a vast variety of multimedia services like voice, email, short messages, Internet access, m-commerce, to mobile video conferencing, streaming video and audio.

Large and specialized databases together with these technological achievements have brought true mobile multimedia experiences to mobile customers. Multimedia implies adoption of new technologies and challenges to operators and infrastructure builders in terms of ensuring fast and reliable services for improving the quality of web information retrieval.

Huge amounts of multimedia data are increasingly available. The knowledge of spatial and/or temporal phenomena becomes critical for many applications, which requires techniques for the processing, analysis, search, mining, and management of multimedia data.

We take here the opportunity to warmly thank all the members of the MMEDIA 2021 technical program committee, as well as all the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and effort to contribute to MMEDIA 2021. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions. We also thank the members of the MMEDIA 2021 organizing committee for their help in handling the logistics of this event.

MMEDIA 2021 Chairs

MMEDIA 2021 Steering Committee

Hiroshi Ishikawa,Tokyo Metropolitan University, Japan

José António Filipe, University Institute of Lisbon - School of Technology and Architecture, Portugal

MMEDIA 2021 Advisory Committee
Jean-Claude Moissinac, TELECOM ParisTech, France
Daniel Thalmann, Nanyang Technological University, Singapore

MMEDIA 2021 Industry/Research Advisory Committee
Giuseppe Amato, CNR-ISTI, Italy

MMEDIA 2021 Publicity Chairs
Lorena Parra, Universitat Politecnica de Valencia, Spain
Jose Luis García, Universitat Politecnica de Valencia, Spain

MMEDIA 2021

Committee

MMEDIA 2021 Steering Committee

Hiroshi Ishikawa, Tokyo Metropolitan University, Japan
José António Filipe, University Institute of Lisbon - School of Technology and Architecture, Portugal

MMEDIA 2021 Advisory Committee

Jean-Claude Moissinac, TELECOM ParisTech, France
Daniel Thalmann, Nanyang Technological University, Singapore

MMEDIA 2021 Industry/Research Advisory Committee

Giuseppe Amato, CNR-ISTI, Italy

MMEDIA 2021 Publicity Chairs

Lorena Parra, Universitat Politecnica de Valencia, Spain
Jose Luis García, Universitat Politecnica de Valencia, Spain

MMEDIA 2021 Technical Program Committee

Manuel Alberto M. Ferreira, University Institute of Lisbon - School of Technology and Architecture, Portugal
Pedro A. Amado Assunção, Instituto de Telecomunicações|Politécnico de Leiria, Portugal
Giuseppe Amato, CNR-ISTI, Italy
José António Filipe, University Institute of Lisbon - School of Technology and Architecture, Portugal
Hugo Barbosa, Lusofona University of Porto / Faculty of Engineering of the University of Porto, Portugal
Letizia Bollini, Free University of Bozen-Bolzano, Italy
Fernando Boronat Seguí, Universitat Politecnica de Valencia-Campus de Gandia, Spain
Dumitru Dan Burdescu, Universityof Craiova, Romania
Baoyang Chen, Central Academy of Fine Arts, Beijing, China
Trista Chen, Inventec Corporation, Taiwan
Minh-Son Dao, National Institute of Information and Communications Technology, Japan
Vincenzo De Angelis, University of Reggio Calabria, Italy
Franca Debole, Institute of Information Science and Technologies - Italian National Research Council (ISTI-CNR), Pisa, Italy
Jana Dittmann, Otto-von-Guericke-University Magdeburg, Germany
Vlastislav Dohnal, Masaryk University, Brno, Czech Republic
Filiz Ersoz, Karabük University, Turkey
Taner Ersoz, Karabük University, Turkey
Tolga Genc, Marmara University, Turkey

Copyright Information

Table of Contents

Design and Evaluation of a Second Screen Interactive Digital Media Solution using MPEG-DASH

Jomar Brudeli
Department of Electronic Systems, IE
NTNU
Oslo, Norway
jomar.brudeli@gmail.com

Shafaq Irshad
Department of Electronic Systems, IE
NTNU
Trondheim, Norway
shafaq.irshad@ntnu.no

Andrew Perkis
Department of Electronic Systems, IE
NTNU
Trondheim, Norway
andrew.perkis@ntnu.no

***Abstract*—This research work designs and evaluates synchronized second screen interactive digital media using Moving Picture Experts Group Dynamic Adaptive Streaming over HTTP (MPEG-DASH). Second screen technology involves the use of a computing device, such as a smartphone or tablet, to provide an enhanced viewing experience for content on another device. The study design uses a structured format for defining MPEG-DASH inline events to trigger the second screen events. Furthermore, the ability for users to make choices that impact the story was implemented by generating several manifests and enabling the trigger manifest to refresh after each question. By performing a formal subjective user experiment, this study assesses the Quality of Experience (QoE) performance of the prototype. The results were positive and strongly indicate a high QoE for the entire system. A secondary result of our work is enhancing the QoE methodology for the assessment of second screen applications. The implemented system can easily be incorporated in the MPEG-DASH standard, allowing support for using the inline event for second screen triggering.**

***Index Terms*—Interactive media; Streaming; MPEG-DASH; Entertainment; QoE; Second screen**

I. Introduction

The world of entertainment and media consumption is constantly changing. It has evolved from stories around the bonfire and painting in caves to cinemas, black and white televisions and finally ended up as the endless amounts of content available on online streaming services today. New mixed media allow a combination of new technologies, such as Second Screen with Interactive storytelling [1].

There has not been a lack of attempts at combining the traditional forms of entertainment with the new technologies [2]–[6], where results have usually just turned out to be some gimmick, adding little to nothing to the overall end-user experience. In the context of 3D movies, however, the recent popularity of Netflix's blockbuster interactive movie "Black Mirror: Bandersnatch" has shown that the audience might be ready for something more [7]. This has led to two of the most prominent players in the industry, i.e., Netflix and YouTube, announcing that they have plans to increase their investments in creating interactive digital media content [8].

The second screen can be a tool for storytelling, interactivity, and increase immersion. A standardized way to utilize all these secondary screens in video streaming technologies is to use Moving Picture Experts Group (MPEG) - Dynamic Adaptive Streaming over HTTP (DASH) [9] [10]. MPEG and 3rd Generation Partnership Project (3GPP) have developed MPEG-DASH to enable interoperability in the industry [11] [12].

One of the typical applications of DASH is audiovisual streaming service due to its high bit-rate. Several studies have been performed to evaluate the Quality of Service (QoS) of audiovisual services [13], where they suggest empirical quality models of audiovisual content [14]. However, there has been very little work on the Quality of Experience (QoE) assessment of such video streaming technologies. We believe it is essential to evaluate the perceived QoE of second screen streaming solutions for finding the optimal balance between available network resources and quality. In this paper

1) We studied a solution for streaming audiovisual content for second screen applications using the MPEG-DASH standard. The goal was to support the viewers by extending and enhancing broadcast TV without disrupting the viewers' attention.
2) We studied using the MPEG-DASH standard to stream audiovisual content for the second screen by designing a second screen streaming solution using MPEG-DASH. The application aimed to support the viewers by extending and enhancing broadcast TV without disrupting the viewers' attention.
3) We presented a solution for synchronized second-device media content that could be used to create a more immersive experience for users. A methodology for QoE testing of interactive second screen prototype is also described. A subjective user experience evaluation was conducted to evaluate the prototype and assess its quality.
4) For the subjective testing, four different event types were used in four versions of the video (in the second screen

 ISBN: 978-1-61208-839-6

prototype) to ensure consistency in the results.

5) Results of the subjective testing were evaluated using two approaches. First, the results of version V1 (event types: none) were compared with V3 (Event type: vibration). In the following approach, V2 (event type: vibration, sound, video, and web) was compared with v4 (event type: sound, video, and web). Mean Opinion Score (MOS) was used to test the subjective performance of the system, and a t-test was used to compare results between the different versions.
6) We studied whether the means of the two distributions are equal. Furthermore, if there is a significant change in the QoE between different versions. (i.e. V1 vs V3 and V2 vs V4)
7) The validation provided us with insights into whether it is possible to develop a method for synchronized second screen content based on existing international standards for video codecs or streaming technology or not.

The paper is organized as follows. Section II defines some literature and theories that are needed to keep up with the method. Section III describes the design methodology in detail, including essential terminologies and specifications of the system. Section IV gives a systematic overview of subjective user evaluation. Sections V and VI discuss the results, and Section VII is a conclusion to the paper summing up the essential points and findings of our work.

II. Background and Related Work

A. Second Screen Technology

Second screen interactive digital media platforms are used nowadays for detecting and dynamically synchronizing media content (e.g., television programs or movies) that a viewer can watch while providing related content on a second screen for enhancing the viewer experience [15].

Using second screen interaction to enhance media consumption experience is not a new idea, and several attempts have been made to enhance the viewers' experience with second screens [16]. In sports, Tour de France developed an app that gave viewers real-time information synced to the live race [4]. In the gaming industry, PlayStation released "Hidden Agenda", a game where friends get together and play the game together on their mobile phones [5]. The dutch movie APP [3] from 2014 came with a supplementary mobile app that used audio to sync to the movie, showing other videos and information.

The latest attempt of creating an interactive cinematic is the movie Late Shift [2] [6] giving the audiences the possibility of changing the course of the movie by making choices and other interactive elements on a synced application. Late Shift even offered these options in the cinema, letting moviegoers vote for the choices that the characters in the movie should make. Current research is evolving around methods and systems to display content playing on the second screen device.

B. Streaming Technology

Dynamic Adaptive Streaming (DASH) over Hypertext Transfer Protocol (HTTP), ISO/IEC 23009-1 [9], commonly-known as MPEG-DASH, is an ISO standardized streaming protocol that enables adaptive bit-rate video streaming over HTTP. This is similar to Apple's HTTP Live Streaming (HLS) and Adobes HTTP Dynamic Streaming (HDS), but not open to all and not bound by any company licensing and limitations. For an extended period, HLS has been the preferred solution for video streaming. However, now we see that the transition to DASH is getting along well with many of the major streaming companies, including Google and Netflix [10].

All modern browsers that support HTML5 and Media Source Extension (MSE) now support MPEG-DASH. The only significant exception from this is Apple's Safari browser, which does not support MSE. There are no reports yet regarding when that might happen.

To offer an adaptive bit-rate for the stream, the content needs to be encoded in different resolutions and then divided into small segments in time. In addition to all these small segments in different bit-rates, the stream would also usually include subtitles and audio in different languages and other metadata. To keep control over all these different parts, MPEG-DASH uses a Media Presentation Description (MPD), also known as a manifest. This is an XML document that follows a specific structure that tells the client where to find all the different parts and modify them. In Figure 1, the high-level representation of an MPD is showed.

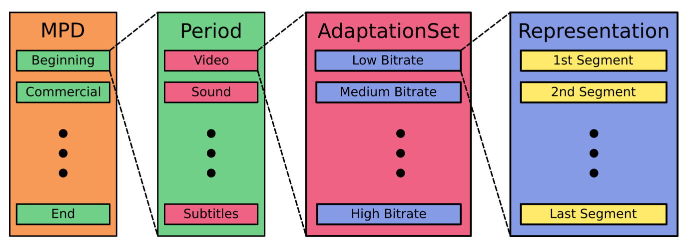

Fig. 1. High level illustration of how a simple MPD could be structured. Figure adapted from [10]

The MPEG-DASH standard ISO/IEC 23009-1 [9] defines that events may be signaled in the MPD on **Period** level. The EventStream element contains two different attributes, @schemeIdUri and @value. The @schemeIdUri provides a Uniform Resource Identifier (URI) to determine which scheme the event follows. The @value attribute is optional and used to distinguish between different types of events in a scheme. In the standard, an MPD validity expiration event is specified as an inline event. The event has the following URI:

```
urn:mpeg:dash:event:2012
```

This research paper uses existing international media standards for the video to develop second screen applications that enhance users' experience. Embedding and extracting the information from and into the media stream is an important aspect. The second screen's content should be triggered synchronized with the original media presentation at specific predefined points or frames.

III. Design method

The method used in this study is based on a user-centered design, and the use case is based on inserting trigger points in a Sci-fi movie developed by the mixed-media production company KapOow [17]. The scene used for the test and validation required trigger points for specific events, for example, opening a breaking news radio segment about 10 seconds into the clip. To investigate the QoE of the interactive second screen, a prototype was developed for testing purposes where different trigger points were implemented.

The prototype was evaluated by a target group, chosen from Sci-fi interested candidates with a keen interest in using mobile technology. The validation criteria were to optimize user QoE. Our definition of QoE is according to the Qualinet White Paper on Definitions of QoE [18]. The definition is: "The degree of delight or annoyance of the user of an application or service. It results from the fulfillment of users' expectations concerning the utility and enjoyment of the application or service in the light of the user's personality and current state."

Using techniques like in-depth interviews and user experiments to get feedback from users regularly during the development of the product, one can make sure that the final solution ends up being something the end-users want to use. Here, User-Centered Design (UCD) is used, a design methodology that focuses on involving end-users in developing a product. This research also investigates how to assess the QoE for end-users of second screen applications by using the prototypes in a subjective evaluation experiment, focusing on the effect of triggering synchronized vibrations on the second screen.

The development of the Android Application and the database structure is not within the main focus of this research paper; however, the detailed documentation and source code can be found at the public GitHub repository [19]. Important terms used are described in Table I.

TABLE I
Table describing important terms

Trigger-event:	Event that occurs on second screen (e.g. video, sound)
Trigger-point:	Timestamp for trigger-event (HH:MM:SS)
Trigger-message:	Defines the trigger-event (e.g. video::filename.mp4)
Trigger-information:	Information containing trigger-points and trigger-message for all trigger-events that are defined for a video.

The next section describes the specifications and a step-by-step procedure of development.

A. Specifications

Trigger information

- solution based on the MPEG-DASH standard (See Fig 3).

The master screen contains

- a custom streaming server.
- a webpage integrating the Dash.js player, also providing login and user identification support.
- some form of login and user identification.

We will also explore the possibility of interactivity (path-choice) using existing technology in the MPEG-DASH standard.

The second screen contains

- functionality for playing sound and video.
- functionality for vibrations.
- a solution to log on to a session based on scanning a QR-code shown on the master screen.
- support for interactivity and voting with several users in the same session.

B. Development

Figure 2 shows the setup for the prototype.

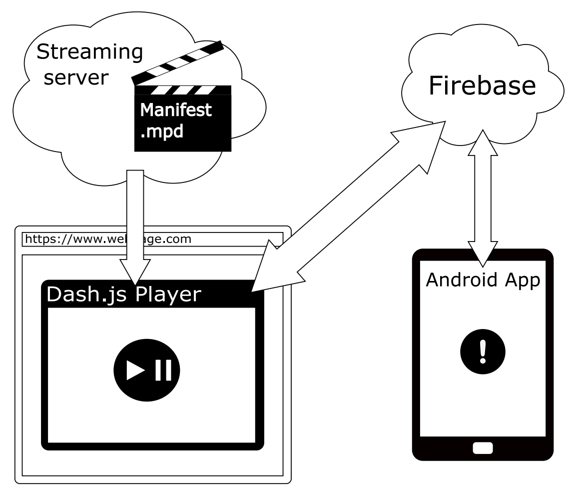

Fig. 2. Simple sketch of the setup used in the prototype

1) Encoding video files: To ensure the best possible experience for streaming, the desired video was encoded into different bit-rates (resolutions) to let the player jump between them depending on available bandwidth, i.e., adaptive bit-rate. This is a common way of doing video streaming today and made the setup as realistic as possible. The test video was already h.264 encoded with a framerate of 25 fps and placed in an MP4 container. Table II shows the five different resolutions and corresponding bit-rates used in the prototype.

The interval between keyframes was set to 25. With a framerate of 25 fps for the original video, this will guarantee

 ISBN: 978-1-61208-839-6

TABLE II
TABLE SHOWING THE RESOLUTIONS AND CORRESPONDING BIT-RATES CHOSEN FOR THE VIDEO. ALL IS IN THE 16:9 ASPECT RATIO.

Resolution[p]	Bit-rate [kbps]
426x240	260
640x360	600
848x480	1060
1280x720	2400
1920x1080	5300

that one keyframe will be located in each segment when the segment length is set to 1 second, after encoding the videos as put into an MP4 container, using MP4Box. The H.264/AVC encoder x264 was used to do the encoding. The details of parameters used can be found in the bash script written for the task, *generate_dash.sh*, found in Appendix A.

2) Generating Manifest: When all the videos have been prepared, they must be divided into segments, and then a manifest must be generated. Using the multimedia packager MP4Box, the manifest can be automatically generated for us by adding the correct parameters. In this project, the following parameters were used:

```
MP4Box -dash 1000 -rap video_1080.mp4
... video_240.mp4 video.mp4#audio
```

The `-dash` option specifies the duration of the segments in milliseconds. A segment duration of 1 second is considered a good value. The `-rap` specifies that every segment should start with a random access point. Finally, the `#audio` modifier extracts the audio track from the source file.

3) Adding Events: For this project, the dash.js [20] reference client implementation for MPEG-DASH playback via JavaScript was used.

To embed the event information in the media stream, it was decided to use inline events. Inline events are declared inside the manifest on the Period level, as shown in Figure 3.

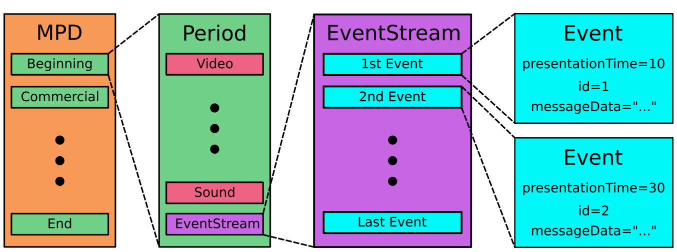

Fig. 3. High-level model showing how EventStreams and Events are placed in a MPD.

Figure 4 shows an example of how to write a functioning EventStream inside the MPD. There are a few important things to notice. First, the `id` of each of the Events must be unique; otherwise, they will not function properly. Secondly, the `presentationTime` property sets the time in second for when an event is triggered. Finally, the `schemeIdUri` property of EventStreams is important for identifying what event this is.

```
<EventStream schemeIdUri="urn:mpeg:dash:event:trigger:2019 value="1">
  <Event presentationTime="10" id="1" messageData="video::test_video.mp4" />
  <Event presentationTime="30" id="2" messageData="sound::test_sound.wav" />
  <Event presentationTime="60" id="3" messageData="vibration::1000,255" />
  <Event presentationTime="90" id="4" messageData="web::https://www.ntnu.no" />
</EventStream>
```

Fig. 4. EventStream that follows our suggested format for declaring trigger-events. All four supported event-types are declared in the example at trigger points 10 sec, 30 sec, 60 sec and 90 sec.

In the MPEG-DASH standard, two Uniform Resource Identifiers (URI) were specified, each with a specified set of rules for what the client should do in response. Dash.js handles the events internally and does not expose them to the developers using the client. This makes it impossible for us to detect an event from outside the client. Also, the standard specifies that events should be deleted after they are triggered, making it impossible to rewind in a video and trigger the event several times.To overcome these complications, the following modifications were made:

- Define a new URI for second screen events.
- Make the client expose the trigger instead of handling it internally.
- Do not remove events after they have been triggered.

To implement our suggested changes, the dash.js had to be modified to suit our needs. Specifically the *Eventcontroller.js*, found in Appendix A.

The standard states that all users may define a schemeIdUri for their respective application, as long as it does not conflict with any other URI in the standard. We suggest the following URI for second screen events.

```
schemeIdUri="uri:mpeg:dash:event:
trigger:2019"
```

Another property worth mentioning is the `value` parameter of the `EventStream` objects. This could be used to distinguish between different EventStreams, e.g., to tailor events to users in different countries or with different devices. For this project, the default value of 1 was used.

4) Path Choice: To let the viewer impact the story they are watching, path choices were implemented using MPEG-DASH. The following trigger-message was suggested.

```
choice::<dur>,<opt_1>,<branch_1>,
    <opt_2>,<branch_2>
```

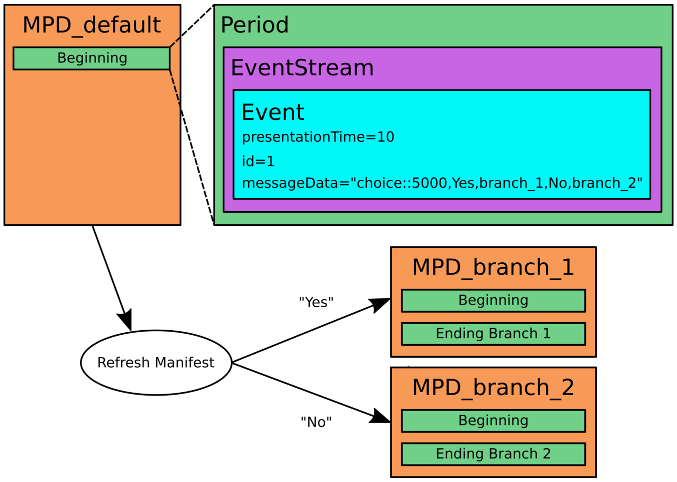

Fig. 5. High-level implementation of path choice using MPEG-DASH standard

Three manifests are uploaded to the webserver (Firebase),

- VIDEO_default.mpd,
- VIDEO_branch_1.mpd
- VIDEO_branch_2.mpd

First, the default manifest is loaded, then the video is initialized. Our examples consist of only one Period that contains the media for the beginning of the video. When a choice-event is triggered on the second screen, two buttons appear, making the user chose between two options, opt_1 and opt_2. A timer is also counting down the time. When the timer has reached zero, a refresh manifest is triggered. If the viewer chose opt_1 then when the player requests a refresh of the manifest, it receives VIDEO_branch_1.mpd and the opposite for opt_2. If none of the options is chosen by the user before the time runs out, a random option is picked. The logic handling the manifest refresh is handled with cloud functions in the webserver. Figure 5 show an example of how this is implemented.

All the functions described for the prototype were verified in the lab, and QoE testing was performed.

IV. Evaluation - Quality of Experience Assessment

A subjective user experiment was conducted to evaluate the prototype and to assess the QoE of the system. The central hypothesis to be tested was that the second screen vibration event positively affected the quality of experience.

KapOow contributed to the video that was used during the experiment. Four different manifests, V1, V2, V3, and V4, were prepared, all with different events included. Table III shows which events were included in which versions.

TABLE III
TABLE SHOWING THE SECOND SCREEN EVENTS ENABLED IN EACH OF THE FOUR VERSIONS OF THE VIDEO USED IN THE SUBJECTIVE USER EXPERIMENT

Version	Events types
V1	None
V2	Vibration, Sound, Video and Web
V3	Vibration
V4	Sound, Video and web

To ensure that there were no systematic errors due to the order in which the participant view the different versions, each participant viewed them in random order.

1) Statistical analysis: To evaluate the effect of vibration on the overall quality of experience, two different approaches are chosen. The first is to compare the version results without any event occurring on the second screen (V1) and the version with only the vibrations events occurring (V3). Next, the version with all events (V2) is compared to V4 (everything except vibration).

The MOS is commonly used in telecommunication to evaluate the subjective performance of a system (e.g. sound quality). The MOS for a specific question is defined as

$$MOS = \frac{\sum_{i=0}^{n} R_i}{n} \quad (1)$$

where R_i is the score from participant i, and n is the total number of participants.

To determine if the MOS for each question in the subjective questionnaire significantly increases when vibration was added. Paired observations Student's t-Test between the score of questions with and without vibration was used. The differences $d_1, d_2, ..., d_n$ between the scores for each of the n questions are used as a basis for the hypothesis test. They are assumed to be drawn from the stochastic variables $D_1, D_2, ..., D_n$, each with mean $\mu_D = \mu_2 - \mu_1$, and variance σ_D. The hypothesis to be tested is given by

$$\begin{aligned} H_0 &: \mu_d = 0 \\ H_1 &: \mu_d \neq 0 \end{aligned} \quad (2)$$

Where H_0 states that the means of the two distributions are equal, and therefore no significant change has occurred. If we can successfully reject H_0 on the other hand, based on our results, we can successfully say that we have an increased QoE.

V. Result and Analysis

Twenty users between 18 and 26 years of age were chosen to participate. As the media provided by KapOow to use in the experiment was in a particular genre (Sci-Fi/ horror/comedy), the participants were chosen within the assumed target audience. 65% of the users were male, and 35% were female. However, even if this genre might be more appealing to a male audience, the gender distribution ended up being quite well distributed.

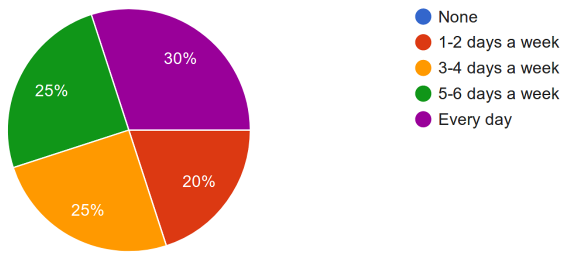

Fig. 6. Figure demonstrating the days of week participants used streaming services (Netflix, HBO, Amazon Prime, etc.) to watch movies and/or series

The streaming habits of the participants were also considered an important parameter worth measuring. While all of them used streaming services weekly, the amount of use was well distributed, as Figure 6 shows.

The MOS for all ten questions was calculated for all four versions. Figures 8 and 9 show how the MOS evolved over the questions. We can observe that the version without any events (V1) occurring on the second screen generally performed worse than the rest. It also shows that the version that showcased all the different types of second screen events performed best (V2).

With only 20 participants in the test, it is limited how much we can deduct from the MOS, at least not without considering the scores' spread. Therefore the MOS was calculated within a 95% confidence interval as well. Figure 8 shows how the score for questions 1 to 5 behaved, and Figure 9 shows the same for questions 6 to 10. The versions are grouped two-and-two such that it is easy to compare versions that differ only regarding if the vibration was present or not.

To statistically prove that vibration has a statistically significant positive effect on the MOS for each question, a Paired Student's t-test was performed on each question. The first test, T1, compared the version with no events (V1) with the version with only vibration (V3).

The second test, T2, compared the version with all event types (V2) with everything except vibrations (V4). The initial null hypothesis is that the MOS of the compared versions is equal. A P-value below the typical value of 0.05 indicated that the difference is significant and that the null hypothesis, therefore, can be rejected.

The results from the t-Tests can be seen in the Tables below. Table IV a) shows the results from the first t-Test and Table IV b) shows the results from the second. The P-values below the chosen threshold of 0.05 are marked with a *. Of all the 10 questions, 6 of them passes the t-Test in both T1 and T2, and only question number 4 fails in both tests.

Some of the participants made a few additional comments, which are listed below.

- "In the end, a bit much vibration in my opinion".
- "Notification/commercial in the middle of experience was like breaking the 3rd wall and pulled me out of the immersion. Sound/vibration and a second screen for news did enhance the experience."

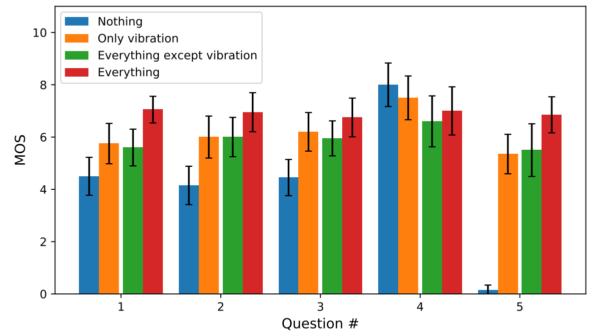

Fig. 8. MOS for questions 1 to 5, with a 95% confidence interval

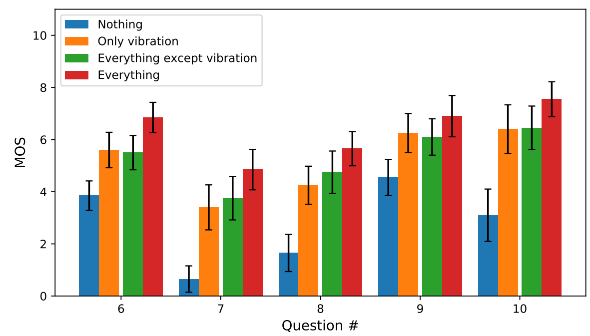

Fig. 9. MOS for questions 6 to 10, with a 95% confidence interval

- "Lack of sync between vibration and sound. Vibrations should be modulated with the sound."
- "Notifications for ads that paused the movie broke the immersion quite a bit. vibration for the garage door lasted a bit long. Other than that: had a very nice experience"
- "Enjoyed the vibration and sound from the synchronized second screen, but the video did not feel natural and paused the story without adding something essential to it. All in all a great experience."
- "The screen in itself seems somewhat distracting, but a media for receiving makes the experiment considerably richer, at least for the horror elements. Could maybe work just as well with a handheld device without a screen (e.g. gaming controller). The light from the second screen was somewhat distracting."
- "It was cool!"

VI. Discussion

For second screen applications to be a successful extension to multimedia content, at least three critical factors should be in place. Technology, content, and audience. In this paper, we have focused on the technology aspect and the audience experience through new ways of consuming stories. The subjective user evaluation experiment results indicated that the technical solution suggested in this work has great potential and that the

TABLE IV
TABLE SHOWS MOS FOR ALL FOUR VERSIONS. THE TABLE ALSO INCLUDES P-VALUES CALCULATED FROM BOTH PAIRED STUDENT'S T-TESTS, T1 AND T2. (* P-VALUE LARGER THAN 0.05. NOT SIGNIFICANT POSITIVE CHANGE IN MOS)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
(a)MOS V1	4.500	4.150	4.450	8.000	0.150	3.850	0.650	1.650	4.550	3.100
MOS V3	5.750	6.000	6.200	7.500	5.350	5.600	3.400	4.250	6.250	6.400
P-value T1	0.056*	0.006	0.008	0.144*	0.000	0.003	0.000	0.000	0.0023	0.000

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
(b) 1 MOS V2	7.050	6.950	6.750	7.000	6.850	6.850	4.850	5.650	6.900	7.550
MOS V4	5.600	6.000	5.950	6.600	5.500	5.500	3.750	4.750	6.100	6.450
P-value T2	0.001	0.031	0.149	0.408*	0.056*	0.002	0.032	0.046	0.104*	0.025

QoE for the system is promising. There are, however, a few remarks that need to be emphasized.

The synchronization is the weakest part of the technical solution. It does not consider the variation in delay for the application to detect a change in the database. During the lab validation, it did not appear to be a problem as long as it took the setting of the trigger times for the events into account. However, a few participants commented on slight delays, indicating that the variation between sessions is too significant to be ignored. This is probably the biggest reason why this technology's most popular commercial implementations have used watermarked audio. However, we think that our solution has several advances over this technology, The most important one being that the user can watch the video with their headsets. It reduces the device's processing power demands, and the users are not dependent on good audio conditions. The synchronization issue should be investigated further, but a simple solution might be implementing time correction/synchronization protocols between the client player and the second screen application. It should also be mentioned that not all event types have the same requirements for synchronization, but vibration and sound most certainly do.

Moreover, is it essential to embed the trigger information in the media file or stream? It might seem like a complicated solution to a simple problem. For this solution, we used the .trig file to store the information before it gets uploaded and parsed into the database and embedding the inline event from MPEG-DASH. By using already implemented functionality from an international standard, the process gets much more manageable. DASH compatible web players already have support for listening after inline events. The only thing developers need to do is specify what event they are waiting for in their JavaScript environment, keeping the complexity as low as possible.

We have suggested and implemented a few different second screen events for this project, including sound, video, path-choice, and web notifications. These are just suggestions developed in collaboration with content creators. There is no limit to what might be triggered on the second screen, as long as the MPEG-DASH client detects the events and the surrounding application implements them. This is an important aspect. Instead, the technology should not be in focus, work as a toolbox for creative minds to utilize for creating exciting and involving content. Not create content and then throw some second screen content on it afterward, which would make the technology just a gimmick.

This is one of the significant problems with the evaluation of the system. The content was created for regular linear storytelling and was just converted into a second screen experience by an engineer. This is not at all ideal and would probably affect the QoE significantly. Throwing some new tech at some old content and just expecting the audiences to embrace it is naive. Netflix's success with "Black Mirror: Bandersnatch" is an excellent example of this. The story and interactivity were tailor-made to fit each other, enhancing the experience. Later attempts have not reached the same popularity. The fact that the content is considered an influential factor is one of the essential differences between QoS and QoE. Even if the system and the technology work perfectly, the users' experience will be damaged if they do not feel that the technology contributes to the story in a good way. Take question 5 from the evaluation experiment, "Do you feel the second screen contributed to the story?". This was the question that showed the most significant jump in MOS between having no events (V1) and having just vibration events(V3), see Table IV. But still only reached a score of 5.35 on the scale "Not at all" (0p) to "Very much" (10p), reaching barely over the middle of the scale.

VII. CONCLUSION

The most important discoveries from this work include using MPEG-DASH inline events to trigger the second screen events. Also, a structured format for defining trigger-information has been suggested. Finally, users' ability to make choices that impact the story was implemented by generating several manifests. The solution was made possible with existing standards and infrastructure and minor changes to MPEG-DASH clients. Evaluation of the prototype was done with a subjective user experiment, and the results were positive

 ISBN: 978-1-61208-839-6

and strongly indicated a high QoE for the entire system. The subjective user experiment was centered around the hypothesis that synchronized vibration events on the second screen would positively affect the QoE. In 6 out of 10 questions, we observed a significant increase in MOS compared to versions without vibration. From this, we can conclude that the hypothesis is, to a certain degree, confirmed. The results can be used in the ongoing development of the DASH standard. With few modifications to the current standard, using the inline event for second screen triggering could be implemented.

ACKNOWLEDGMENT

A big thank you goes to Patrick Velasco for developing both the prototypes.

REFERENCES

[1] H. Simon, E. Comunello, and A. Von Wangenheim, "Enrichment of interactive digital tv using second screen," *International Journal of Computer Applications*, vol. 64, no. 22, pp. 58–64, 2013.
[2] (2016) Late shift - the world's first cinematic interactive movie. [retrieved: March, 2021]. [Online]. Available: http://lateshift-movie.com/
[3] (2014) Review: 'app' is the first movie designed for you to use your phone while you watch, and it's not bad. [retrieved: March, 2021]. [Online]. Available: https://www.indiewire.com
[4] (2018) Letourdefrance: iphone, ipad and android official apps. [retrieved: March, 2021]. [Online]. Available: https://www.letour.fr/en/mobiles-apps
[5] (2019) Hidden agenda — ps4 games — playstation. [retrieved: May, 2019]. [Online]. Available: https://www.playstation.com/en-au/games/hidden-agenda-ps4/
[6] (2019) Ctrlmovie: An audience-collaborative cinematic experience. [Online]. Available: http://www.ctrlmovie.com/
[7] E. Pajorová and L. Hluchý, "Augmented reality as a higher education form for students with delimited ability," in *Smart Education and e-Learning 2019*, V. L. Uskov, R. J. Howlett, and L. C. Jain, Eds. Singapore: Springer Singapore, 2019, pp. 461–469.
[8] L. Shaw. (2019) Youtube is developing choose-your-own-adventure programs. [retrieved: March, 2021]. [Online]. Available: https://www.bloomberg.com/news/articles/2019-04-09/youtube-is-developing-choose-your-own-adventure-programs
[9] "Information technology – dynamic adaptive streaming over http (dash) – part 1: Media presentation description and segment formats," International Organization for Standardization, Standard, 2014.
[10] (2019) Mpeg-dash (dynamic adaptive streaming over http, iso/iec 23009-1). [retrieved: March, 2021]. [Online]. Available: https://bitmovin.com/dynamic-adaptive-streaming-http-mpeg-dash/
[11] T. Stockhammer, "Dynamic adaptive streaming over http–: standards and design principles," in *Proceedings of the second annual ACM conference on Multimedia systems*. ACM, 2011, pp. 133–144.
[12] T. C. Thang, Q.-D. Ho, J. W. Kang, and A. T. Pham, "Adaptive streaming of audiovisual content using mpeg dash," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 1, pp. 78–85, 2012.
[13] S. Takenaka, K. Kanai, and J. Katto, "Qos improvement of mobile 4k video by using radio quality map," in *2015 IEEE 4th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2015, pp. 541–542.
[14] T. C. Thang, J. W. Kang, and Y. M. Ro, "Graph-based perceptual quality model for audiovisual contents," in *2007 IEEE International Conference on Multimedia and Expo*. IEEE, 2007, pp. 312–315.
[15] B. Slavin, S. Rosenberg, and A. Glennon, "Second screen interactive platform," May 2 2013, uS Patent App. 13/621,277.
[16] K. Blumenstein and et al., "Visualizing spatial and time-oriented data in a second screen application," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '17. New York, NY, USA: ACM, 2017, pp. 84:1–84:8. [Online]. Available: http://doi.acm.org/10.1145/3098279.3122127
[17] (2019) The next generation mixed-media production company. [retrieved: March, 2021]. [Online]. Available: https://kapoow.no/
[18] K. Brunnström, S. A. Beker, K. De Moor, A. Dooms, S. Egger, and e. a. Garcia, "Qualinet white paper on definitions of quality of experience," 2013.
[19] J. Brudeli. (2019) Github repository, they came. [Online]. Available: https://github.com/jbrudeli/they-came-colab
[20] (2019) dash.js. [retrieved: March, 2021]. [Online]. Available: https://www.videolan.org/vlc/index.html

APPENDIX A
APPENDIX (SOURCE CODE)

A. GitHub repository

https://github.com/jbrudeli/they-came-colab

B. Statistical analysis

https://www.dropbox.com/s/blicxhdz9d48zxi/analysis.py?dl=0

C. Generate DASH Manuscript, bash script

https://www.dropbox.com/s/i9p8qlfon1t6j9q/generate$_dash.sh?dl=$ 0

D. Modified dash.js Eventcontroller

https://www.dropbox.com/s/9hkv67qf7yaoyeo/EventController.js?dl=0

Human Emotion and Machine Emotion - Studies of Emotion in AI

Wang Shuo
School of New Media and Art Design, Beihang University
Beijing, China
email: wangshuo_ws@buaa.edu.cn

Yang Shuo
Art Criticism, Chinese National Academy of Arts
Beijing, China
email: flyingonbird@126.com

***Abstract*—Artificial Intelligence (AI) is undoubtedly a hot word in the field of contemporary art in recent years. The consciousness, intuition and emotion of AI have attracted the attention of artists in particular, and many art works have explored this topic. Does AI have the emotional characteristics of humans? Is the emotion of AI equal to the emotion that defines human beings? This paper attempts to step out of the anthropocentric perspective, examine the boundary and relationship between human emotion and machine emotion, and inject a new theoretical perspective into the AI artistic practice. This paper first discusses what emotion is in the biological sense, analyzes whether a machine can have emotion from both positive and negative aspects, and puts forward the definition of "machine emotion". This paper also reviews some theories and practices related to emotion in the AI field. The conclusion of this paper is that although artificial intelligence cannot possess human emotions, it can possess "machine emotions" beyond the narrow sense of human emotions, and the construction of emotional mechanisms inside AI may become a new research direction in this field.**

Keywords-Emotion; Artificial Intelligence; Machine Emotion.

I. INTRODUCTION

Picard [27] told the story that children think Barney the toy dinosaur produces emotions, and Barney's conscious expression encourages them to believe that Barney has emotions. As far as we know, children have always attached their emotions to dolls and stuffed animals, and we all agree with the experience that toys never have and cannot have emotions. So, drawing an analogy between a toy dinosaur and Artificial Intelligence (AI), we have to ask: why would humans think about whether AI can have emotions?

Arkin [3] believes that roboticists mostly focus on pragmatic functionality and ignore emotional features. As humans demand more of AI's ability to communicate, researchers are increasingly assigning some of the characteristics of human emotions to machines. So, are the emotional features of the AI (as Barney the dinosaur) merely empathic effects of human beings projecting their own emotions onto the machine? However, the emotional interaction between AIs and humans goes far beyond what humans and toys can do, and AI's emotional capacity brings to mind the term "machine emotion". At this point, it seems necessary to re-examine the definition of human emotion. Is the definition of emotion based solely on anthropocentrism insufficient to include machine emotion? Can we go beyond human or biological emotions to redefine machine emotions? These are the questions that this paper wants to discuss.

In Section 1, we discuss the emotion in the general biological sense from the aspects of its origin, definition and classification. In Section 2, we argue the motive and purpose of humans giving emotion to AIs, and then analyze whether AIs can have emotion from both positive and negative aspects, and put forward the definition of "machine emotion". In Section 3, we review human-based and machine-based emotions ontologies. In Section 4, by reviewing relevant literatures in recent years, we find that more and more researchers are trying to explore the construction of "machine emotion" inside artificial intelligence machines. The last section is the conclusion of this paper.

II. WHAT IS EMOTION

From a neurobiological point of view, the source of emotion lies in the tight interaction between the amygdala and the cerebral cortex [2]. We know that the emotional expression of organisms is directly related to the transmission of facial expressions, which belongs to the emotion within organisms. In brain movement, the observation of facial expressions can activate the amygdala. If the amygdala is damaged or defective, it will lead to confusion of emotional expression, making it difficult to observe and recognize facial expressions [31]. The amygdala and the prefrontal cortex of the brain are also involved in the construction of social emotions, which are key components of social interaction [10]. The amygdala, for example, is the mechanism by which organisms make judgments about sensory events by triggering the system of expectation, punishment or evaluation associated with emotion, in response to neural stimulation.

Different scholars define emotion from different angles. Mayers [24] defined emotion by relationship between emotion and body: "Emotions are our body's adaptive response". R. Lazarus and B. Lazarus [21] laid emphasis on how emotion reacts to external environment: "Emotions are organized psycho-physiological reactions to news about ongoing relationships with the environment". Plutchik [29] deemed that emotions are involved in all parts of the link from psychological to actual actions, a continuous status: "Emotion is a complex chain of loosely connected events which begins with a stimulus and includes feelings, psychological changes, impulses to action and specific, goal-directed behavior".

Nowadays, there are two major emotion classification models. One is basic emotion theory, and the other is emotion dimension theory. In the context of psychological science, the concept of basic emotions is deemed as a classification that defines a group of similar emotions with specific color. Emotions are marked by extents such as joy, happiness, or ecstasy [19]. Ekman proposed that the "basic" term has three layers of indication: first, the perspective of basic emotions could help us to distinguish those perspectives that believe emotions are consistent and only differs in intensity and extent of joy, such as some simple classifications that counterpose negative and positive emotions; second, basic emotional states reflect the fact that

 ISBN: 978-1-61208-839-6

emotions are of adaptive value evolution when dealing with basic life tasks. It is embodied in human's status when they handle repeated tasks (such as, response to interpersonal relationship, response to emotional reaction to personal life. Basic emotional states are also related to our intuitive characters and life experience); third, basic emotions are also used to describe components that constitute complex emotions. For instance, smugness might be considered to be a blend of the two elemental *emotions, happiness and contempt* [12].

Many scholars have tried to classify basic emotions (for some attempts in this direction, see [4][11][17][25][28] [38]-[41]. For example, facial emotions of human, according to Ekman's model, have six identifiable basic emotion states, which are, respectively: Anger, Fear, Distress/Sadness, Enjoyment/Happiness, Disgust, and Surprise [11]. Silvan Tomkins [38]-[41] describes innate emotions that are exhibited by toddlers since birth as affects: anger, embarrassment, sorrow, fear, interest, happiness, surprise, disgust. In Minsky's six-element model, it is believed that basic emotions include empathy, envy, love, aggression, awe, respect [25].

In psychological science, there is another classification model called the dimensional emotion, which is frequently utilized in recent emotion measurement research. The dimensional emotion uses some indicators to measure emotional intensity - the extent of how behavior is affected by emotions, such as stimulation, energy, and activation [19]. The first scholar who proposed emotion dimension theory is Russell. He believed that emotions could be indicated and measured by two dimensions of a circumplex: 1) valence, which indicates extent of joy when an individual experienced the state; 2) arousal, which indicates possibility of the individual to take certain action due to its specific state [18].

Afterwards, many scholars conducted their development and exploration about dimension models of emotions [23] [28]-[30][36][37]. Thereinto, Thayer's circle of emotion, Plutchik emotional wheel and Lövheim's cube of emotions are regularly used in recent research [13][32][33][43].

III. WHETHER MACHINES CAN HAVE EMOTIONS

A. Why do humans endow machines emotions?

Why do humans endow machine with emotions? Picard [27] identifies four motivations for human beings to endow machines with certain emotional abilities: 1) The first goal is to build robots and synthetic characters that can emulate living humans and animals – for example, to build a humanoid robot; 2) The second goal is to make machines that are intelligent, even though it is also impossible to find a widely accepted definition of machine intelligence; 3) The third goal is to try to understand human emotions by modeling them; 4) The fourth goal is to make machines less frustrating to interact with. All four of Picard's motivations seem to be human-centric: hoping machines become more human-like in appearance, intelligence, and behavior, and further exploring the intellectual pedigree of human emotions through machines.

Arbib and Fellous [2], on the other hand, point out four reasons why people are interested in giving emotions to machines from the perspective of anthropocentrism: 1) the current technology already shows the value of providing robots with 'emotional' expressions (e.g., computer tutors) and bodily postures (e.g., robot pets) to facilitate human–computer interaction; 2) They raise the question that the value of robots in the future may not just be to simulate human emotional expression, but to actually "have emotions"; 3) It, in turn, requires us to revisit the neurobiology of emotion to generalize some new concepts, because what we know was first proposed for humans and then extended to organisms and machines, which makes it interesting to study emotions in robots; 4) It suggests that building "emotional robots" can also provide a novel test-bed for theories of biological emotion. Arbib and Fellous posed a series of related questions: Will machines in the future actually "have emotions", in addition to imitate human emotional expression? Will theories and concepts based on human beings (e.g., neurobiology, biological emotion theory) lead to new research directions with the birth of robots? This article can't help but ask a further question: how to define "having emotions"? Can robots be included in the category of "living things" like humans and animals? Does "having emotions" mean that robots produce emotions, or even thoughts, that are independent of humans?

B. Machines cannot have emotions

Opponents often argue that machines cannot have emotions in three ways: First of all, the emotional transmission and recognition system of the machine is related to its ability to comprehend, recognize and execute tasks. However, when the designers put too much emphasis on the technical execution of the machine in terms of communication and cognition, this shaping method exposes certain emotional defects. As stated by Parisi and Petrosino [26], the actions of existing "emotional robots" are controlled by symbolic, role-based systems.

Second, AI is based on, but also limited to, human emotions. Becker [5] 's research explores the passive role of machines in human-computer interaction. According to him, machine's emotion is positioned only in its ability performing in activities of social networks; AI entities are artificial objects constructed according to the model of human communication and cognitive ability. They have no personality and are relegated to being an interaction partners of humans.

For example, the Embodied Conversational Agents (ECAs) project attempted to build embodied emotional entities of AIs. In order to enable machines to generate special activities related to human emotions and to regenerate new activities reconstructed by the machine itself, ECAs classifies human emotions to construct samples for emotion recognition. Through psychological experiments, they set up seven basic emotions, including "angry, sad, happy, frightened, ashamed, proud, despairing" [34]. However, Becker [5] believes that these patterns of subjective feelings and the variety of associated physiological processes are only taken into account in a highly reductionist manner. On the one hand, ECAs takes human beings as the basic reference model for the emotional entities of AI, which is an extension of human emotion based on machines. On the other hand, it also reflects the human-centered perspective in the process of emotion observation. Therefore, machine emotions, which are set as samples, have regular processing patterns that limit the scope of machine emotions. Becker [5] believes that when a human makes an observation based on his perspective and interest, he can get the specific emotion he wants to collect. This is a perspective from the observer, which implements a certain dominance for machine emotion. Then, the "emotion" generated by the emotional robot is derived from the designer's understanding and construction of the human emotion model. The ability of machines to perform

 ISBN: 978-1-61208-839-6

emotional tasks contributes to the misconception and illusion that machines have emotions.

Finally, from the point of view of motivation selection, the machine is not capable of producing emotion. The implementation of behavior is directly related to biological motivation. When organisms face competition among various motivations, they will make a choice and trigger strategic emotions.

Arbib's Schema Theory reveals the generation process of various biological motivations and analyzes the strategic emotions caused by biological selections of behaviors. With the praying mantis as a model, he explained that the actions of organisms are not only affected by the changes of external environment, but also related to the visual stimulus situation in which the praying mantis is located, the internal variables of the living body, and the existing experience about the stimulus [5]. Compared with motivation, emotion can provide organisms a more efficient, accurate and rapid channel for their actions [5]. Emotions arise from biological body/brain sensations that present a special state when a creature makes a choice among motivations [5]. For example, feeding and mating are competitive behavioral motivations among different activities of organism. Emotional states, then, combine the creature's perception of the external environment to help it make choices and act on them.

Fellous and Arbib [14] argued that machines could not be considered as having emotions at this point, because they could not control their own behavior to make motivational choices in a given situation. Moreover, in the construction system of AI, motivation is the driving mechanism of the machine and the main manifestation of functionality. The motivation control of artificial intelligence lies with the designer and its communication user. The robot only realizes the technical action, and there is no strategic or motivational response action. As Parisi and Petrosino [26] put it, if a robot does not have autonomous motivation, it will not have emotions.

C. Machine emotion

Arbib and Fellous [2] divide the application of emotion in AI into two parts: one is the external aspect of emotion, which refers to the emotional expression made for communication and social collaboration; the other is the inner aspect of emotion, which can influence behaviors (such as action selection, attention and learning).

Researches concerning the external (or social) aspects of emotion, following the anthropomorphic tendencies, promote the production of robots with emotional and empathic relationships between robots and human beings.

Researches concerning the intrinsic (or individual) aspects of emotion focus on the production of robots' subjectivity whose behavior is influenced by endogenous regulatory processes modelled on natural emotion regulation mechanisms [8]. Parisi and Petrosino [26] point out that "current emotional robots can express emotions or recognize our emotional expressions, but they cannot be considered to have emotions because emotions do not play any functional role in their behavior". According to them, AI that simulates or identifies human expressions of emotion is not considered to "have emotion", but only if it interferes with and influences behaviors. In other words, the inner aspect of emotion more closely meets the criteria that "AIs have emotion".

Giving a robot a mechanism of subjective emotional regulation is considered to be an attempt to truly generate "artificial emotion" [8]. In sum, machine emotion is not the same as human emotion. However, machine can form an autonomous regulation mechanism just like human beings, which can play an important role in influencing its behaviors.

IV. Theories about Emotions and Machine

A. Picard's 4 emotional components

Picard [27] believes that the four emotional components of human beings could be a part of machines, thus, to help machines to better adapt to human beings. The components include emotional appearance, multiple levels of emotion generation, emotional experience and mind-body interaction. The first component she proposed, emotional appearance, includes some actions and behavior with emotional appearance expressed by systems. Machines could simulate facial expression, voice, and physical language of human beings to express emotional appearance similar with the emotion expression of human beings. Such emotional appearance is based on learning by machines of human beings' emotion appearance data, thus enabling the establishment of communication mechanism with human's emotions.

The second component is "multiple levels of emotion generation": human's emotions are of different levels. Different external stimulation triggers different levels of reactions. Machines could use this rule to learn about emotions of human beings, or synthesize an internal emotional status tag for themselves to trigger appropriate reactions. Emotion levels include quick-response sub-consciousness, moderate- response and acquired pre-consciousness, and slow-response reaction generated by rationality [27].

The third component is emotional experience, which actually relates to AI's cognition about the emotional field of its own. Emotional experience refers to the generation by machines of emotional status and series of feelings similar with those sensed by human being. Moreover, machines could also sense the thing their bodies are doing [27]. Though "human feelings" is definitely different from "machine feelings", machines could use such feelings and experience to adjust and improve their behavior.

The fourth component is mind-body interaction: emotions include both physical system changes outside and inside the brain. The interaction between emotions, bodies and cognition status is very active. For example, when someone's required to express love, his behavior of telling truth or a lie is totally different. In other words, emotion expression of his body is selectively interfered by status of telling truth or lie. "If a machine wants to copy human being's feelings, the extent of such copy must involve components like signals and adjustment of such emotions. They constitute interactive connection between the body and mental status" [27].

B. Somatic theory and appraisal

In psychology science, there are two major theoretical factions in discussion about emotions: somatic theory and appraisal. The somatic theory fraction believes that emotions are prior to the cognitive processes: "Before analyzing a sensed entity or even recording of any impressions, the brain could instantly call for emotions related to such entity" [19]. Cañamero [7] proposed that emotions (at least part of emotions) are mechanisms used by biological factors against the environment. This makes generation of autonomy and adaptation easier. Bellman [6] agreed with Cañamero's opinions. He believes that animals with emotions could survive better than animals without emotions. Cañamero [7] proposed that emotions' function in this aspect could be applied to design an autonomous robot. Emotions could enable robots to: 1) react faster; 2) make better decision about choosing among several options; 3) send signals about important events to others.

The appraisal, on the contrary, believes that analysis on process of cognizing stimulation is prior to generation of emotions. Many scholars who research on machine emotions build their emotion model with appraisal as the starting point. Some researchers do their consideration from the actual aspect of engineering. They deem that the function of emotions in cognition (see Figure 1) could strengthen the robots' expression in the establishment of emotion models [7][15]. Furthermore, it has been proven by neurological research that emotions' influence to human beings does not interfere with rationality. On the contrary, emotions are of critical function to some basic rational behavior in our general understanding, such as perception, learning, attention, memory and some other abilities [1][9][22]. Gadanho and Hallam [16] deem that emotions could also affect several basic cognition mechanisms like perception, attention, memory and reasoning.

Gadanho and Hallam [16] further proposed that some features of emotions could be transferred to AIs:

a) Attention control: emotions could, by affecting perception and reasoning mechanisms, force the subject to focus on the most urgent problems;

b) Adaptability strengthening of the subject: emotions could check on behavior of the subject, thus to change plans and actions of the subject when necessary;

c) Memory filtration: emotions could better call up memories consistent with current emotion status. Such memory could help the subject to learn about happiness or sadness they once experienced, thus influence the final decision.

d) Reasoning auxiliary: the actor's emotion system could rapidly acquire perceptive clues that could be used to guide acquisition of cognitive information, so as to support thinking of the cognition system.

e) Behavior trend related to certain emotional scenes or even rigid response: these built-in responses enable automatic triggering of appropriate behavior under urgent circumstances, and thus avoid spending unnecessary time on complicated reasoning.

f) Physiological activation of body: intense emotions are usually related to energy release of anticipated necessary action response. Its application into AI system could include adjustment to system parameters, such as speed of behavior actions;

g) Support to social interactions: emotion expression enable individuals to deliver information essential to their own survival to others, thus gives it high value of adaptability.

It is deemed by many scholars that emotions could affect cognition process, especially in two aspects of issue-solving and decision-making [9] [26] [44]. For them, the most basic and common function of emotions is to "realize more efficient operation of the intentional decision-making mechanism of the body" [26].

Cognition is a periodical behavior that could be classified into several levels. For example, Sloman [35] proposed that cognitive behavior of robots include three major stages: 1) reactive (direct actions); 2) review stage (selecting better and more effective behavior among optional behaviors); 3) integrated management (allowing monitoring, classification, evaluation and control on internal stages). Moreover, Ortony also analyzed mutual influence, motivation (action trend) and cognition (definition) of emotions, as well as three-stage behavior in information processing, respectively are reactive (electronically set action mode- primary level emotions), daily (practiced automatic behaviors- primitive emotions), and reflex (senior cognitive functions, comprehensive cognition, non- consciousness, self-reflected- high level emotions). If third-stage cognitive capability is realized, robots would be required to be capable of handling tasks without preset rules in unpredictable environment. Such machine could be enabled of curiosity and self-awareness, so as to get better ability in problem-solving.

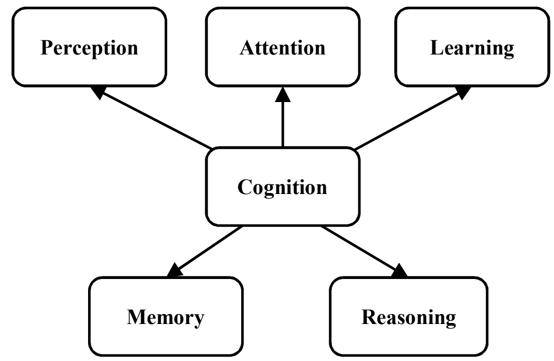

Figure 1. Five aspects of how emotions affect cognition

V. THE STUDIES OF EMOTION IN AI AREA

As mentioned earlier, machines that simulate or recognize human emotional expressions are not considered to have emotions, but are considered to have "machine emotions" only when a machine, as human beings, generates an autonomic regulatory mechanism that plays a role in influencing the machine's behavior planning. It is found that, although many studies of emotion in AI area still focusing on making AI stimulate or recognize human emotional expressions, more and more attempts and explorations have been made in recent years to construct "machine emotion" inside AI machines.

Parisi and Petrosino [26] built an emotional circuit with a neural network to control the behavior of the AI. This mechanism would allow the AI to make more accurate motivational choices. The neural network can break through the rules of motivation and behavior set by human beings and activate the robot's motivational decision-making and emotional representation in response to the situation. This emotional state, dominated by virtual dialogue, can realize machine function in different ways of, from executing aimlessly to behaving under multiple motives. In addition, thanks to AI's computational and mechanical mechanism, it can disassemble the original motive selection of human beings or other creatures with the help of algorithm. According to evaluation theory, among several competing motives or emotions, AI can evaluate the next action that human beings or other creatures are about to take. This decision is made by the AI, which will react to the motives or emotions and affect the future actions of the living creature.

In addition, some studies try to build a bridge between physiological emotions and computer computation. They try to reflect changes of some neuro-modulating substances in human bodies through number changes of computer computation, so as to present own emotion changes of

 ISBN: 978-1-61208-839-6

machines. For example, Kugurakova, Talanov, and Manakhov [20] tried to build a human-simulated subject which has internal emotional statuses, can make responses to emotional stimulation and may show emotions such as sympathy or infringement towards the one it talks with. As believed by them, the real AI feelings need a complicated structure which can decide responses and emotional states of a machine subject. Based on the Emotion Rubik's Cube theory of Lövheim and the chemical and physiological process mechanisms in human brains, they simulated dopamine, serotonin and noradrenaline in a machine model and manifested machine emotions through changes of these neuro-modulating substances. Vallverdú, Talanov, Distefano, Mazzara, Tchitchigin, and Nurgaliev [43] proposed a similar model: Neuromodulating Cognitive Architecture (NEUCOGAR). NEUCOGAR, "based on the architecture of Von Neuman, aims to recognize the mapping from influences of serotonin, dopamine and noradrenaline to computer programs, so as to realize emotion phenomena which can be operated in a Turning machine model". Different from the research of Kugurakova, Talanov, and Manakhov, they expanded the Emotion Rubik's Cube of Lövheim using indexes of the architecture of Von Neuman.

Besides reflection of machine emotion changes based on neuro-modulating substance changes in human bodies, some scholars tried to use some tools, such as variable fuzzy sets and fuzzy cognitive maps, to compute original data, reflect changes in machine emotions, predict these emotions, design an emotion decision making system, etc. For example, Fan Deng, Su, and Cheng [13] presented a prediction model of machine emotions based on emotional dimensions and the theory of variable fuzzy sets. As found in the research, any original data input can be computed by a variable fuzzy set, which provides a mathematical method to express emotion changes which are quantitative, gradually qualitative and mutation qualitative. For another instance, in the thesis published by Salmeron [32], based on the Thayer's emotion model and Fuzzy Cognitive Maps (FCMs), a new method was proposed for prediction of machine emotions and design of an emotion decision making system. As found in his research, machine emotions can be predicted by original data generated by a FCMs sensor. Based on this paper, Salmeron [33] proposed again in 2015: taking Fuzzy Grey Cognitive Maps (FGCMs) as an effective tool to predict machine emotions of an autonomous system immersed in a highly uncertain and complicated environment.

VI. Conclusion

As early as 1950s, Alan Turing already asked the question "whether a machine can think" [42]. A lot of researchers have explored this topic, while "emotion" has always been mentioned in particular as the most outstanding feature that can distinguishes humans from machines, and as believed, perceptual emotions can play a crucial role in rational behaviors. Therefore, scholars mainly focus on the problem whether a machine can have emotions just like humans or not.

First of all, this paper describes the definition of emotions in the general biological sense and tries to explore whether an AI machine can own such emotion. As found in the research, humans give machines the emotions mainly because they want to make machine behaviors further similar with those of humans and then machines can serve humans better. Because of the human centered standpoint, machines cannot possess an independent emotion mechanism. Scholars argue that the expression of simulating or recognizing human emotions cannot be deemed as possession of emotions, so they further propose the concept of "machine emotions" – it can form an automatic regulation mechanism in machines just like human emotions and can influence behavioral organization.

Based on reviewing some theoretical and practical application of emotions in AI during recent years, the research finds that emotion cognition, emotion prediction, emotion-aided decision making or the like are still core topics in the field of emotion and machine research (see Figure 2). However, it is also found in this paper that this field has a new trend to build a machine emotion mechanism inside AI machines. These new studies try to break the borders between creatures and machines, and build a bridge between them. They try to simulate humans' emotion change mechanism in machines rather than implant human emotions in the form of inputting, so that influences of external surroundings on machine emotions are further emphasized.

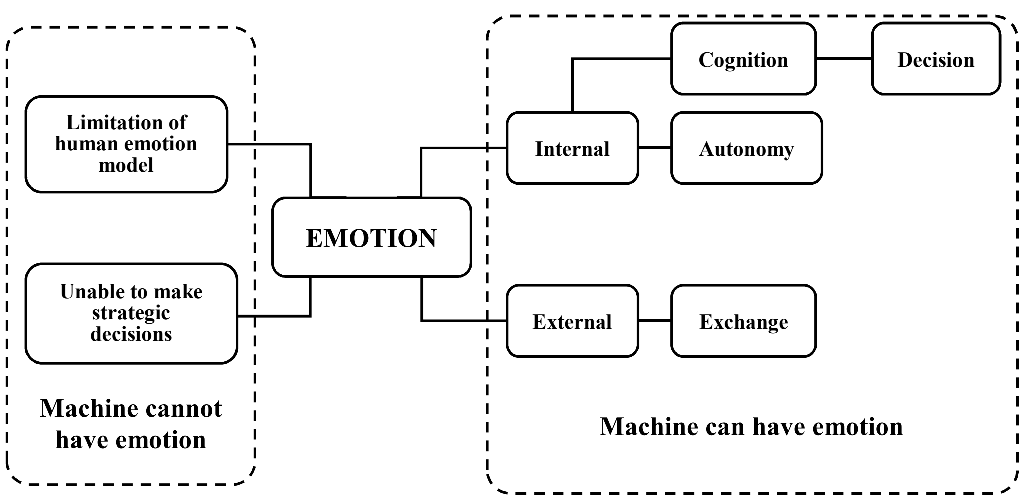

Figure 2. Conclusion

REFERENCES

[1] R. Adolphs, "Is the human amygdala specialized for processing social information?" Annals of the New York Academy of Sciences, vol. 985, No. 1, pp. 326–340, 2003.

[2] M. A. Arbib and J.-M. Fellous, "Emotions: from brain to robot," Trends in cognitive sciences, vol. 8, no. 12, pp. 554–561, 2004.

[3] R. C. Arkin, "Moving up the food chain: Motivation and Emotion in behavior-based robots," in Who needs emotions?: The brain meets the robot. J.-M. Fellous & M. A. Arbib, Eds. New York: Oxford University, pp. 245–269, 2005.

[4] M. B. Arnold, Emotion and Personality. New York: Columbia University Press, 1960.

[5] B. Becker, "Social robots-emotional agents: Some remarks on naturalizing man-machine interaction," International review of information ethics, vol. 6, no. 12, pp. 37–45, 2006.

[6] K. L. Bellman, "Emotions: Meaningful mappings between the individual and its world," in Emotions in Humans and Artifacts. R. Trappl, P. Petta, and S. Payr Eds. Cambridge: The MIT Press, pp. 149–188, 2003.

[7] L. D. Cañamero, "Designing emotions for activity selection in autonomous agents," in Emotions in Humans and Artifacts. R. Trappl, P. Petta, and S. Payr Eds. Cambridge: The MIT Press, pp. 115–148, 2003.

[8] L. Damiano, P. Dumouchel, and H. Lehmann, "Towards human–robot affective co-evolution overcoming oppositions in constructing emotions and empathy," International journal of social robotics, vol. 7, no. 1, pp. 7–18, 2015.

[9] A. R. Damasio, Descartes' Error: Emotion, Reason and the Human Brain. New York: Putnam, 1994.

[10] R. J. Davidson, "Darwin and the Neural Bases of Emotion and Affective Style," Annals of the New York academy of sciences, vol.1000, pp. 316–336, 2010.

[11] P. Ekman, "An argument for basic emotions," Cognition and emotion, vol. 6, no. 3/4, pp. 169–200, 1992.

[12] P. Ekman, "Basic emotions," in Handbook of Cognition and Emotion. New York: Wiley, pp. 45–60, 1999.

[13] K. -K. Fan, S. -Y. Deng, C. -H. Su, and F. -Y. Cheng, "Theory of variable fuzzy sets for artificial emotions prediction," Mathematical problems in engineering, vol. 2015, 2015.

[14] J.-M. Fellous and M. A. Arbib, Eds. Who needs emotions?: The brain meets the robot. New York: Oxford University Press, 2005.

[15] N. H. Frijda and J. Swagerman, "Can computers feel? Theory and design of an emotional system," Cognition and emotion, vol.1, No. 3, pp. 235-257, 1987.

[16] S. C. Gadanho and J. Hallam, "Robot learning driven by emotions," Adaptive behavior, vol. 9, No. 1, pp. 42–64, 2001.

[17] C. E. Izard, Human Emotions. NewYork: Springer, 1977.

[18] A. R. James, "A circumplex model of affect," Journal of personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980.

[19] Z. Kowalczuk and M. Czubenko. "Computational approaches to modeling artificial emotion–an overview of the proposed solutions," Frontiers in robotics and AI, vol. 3, pp. 1–12, 2016.

[20] V. Kugurakova, M. Talanov, N. Manakhov, and D. Ivanov, "Anthropomorphic artificial social agent with simulated emotions and its implementation," Procedia computer science, vol. 71, pp. 112–118, 2015.

[21] R. S. Lazarus and B. N. Lazarus. Passion and Reason: Making sense of our emotions. NewYork: Oxford University Press, 1996.

[22] J. E. LeDoux, "Emotion circuits in the brain," Annual review of neuroscience, vol. 23, No. 1, pp. 155–184, 2000.

[23] H. Lövheim, "A new three-dimensional model for emotions and monoamine neurotransmitters," Medical hypotheses, vol. 78, no. 2, pp. 341–348, 2012.

[24] D. J. Mayers, Eds. "Emotions, stress, and health," in Psychology. NewYork: Worth Publishers, 2010.

[25] M. Minsky, The Emotion Machine and the Society of Mind. NY: Simon and Schuster, 2006.

[26] D. Parisi and G. Petrosino, "Robots that have emotions." Adaptive behavior, vol. 18, no.6, pp. 453–469, 2010.

[27] R. W. Picard, "What does it mean for a computer to have emotions? " in Emotions in humans and artifacts, R. Trappl, P. Petta and S. Payr, Eds. Cambridge: The MIT Press, pp. 213–235, 2003.

[28] R. Plutchik, "A general psychoevolutionary theory of emotion," in Emotion: theory, research, and experience, vol. 1, R. Plutchik and H. Kellerman, Eds. New York: Academic, pp. 3–33, 1980.

[29] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," American scientist, vol. 89, no. 4, pp. 344–350, 2001.

[30] J. Posner, J. A. Russell, and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," Development and psychopathology, vol. 17, no. 3, pp.715–734, 2005.

[31] E. T. Rolls and S. M. Stringer, "Neurophysiology of the primate hippocampus leading to a model of its functions in episodic and spatial memory," 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 2004, pp. 631–636, doi: 10.1109/IJCNN.2004.1379988.

[32] J. L. Salmeron, "Fuzzy cognitive maps for artificial emotions forecasting," Applied soft computing, vol. 12, no. 12, pp. 3704–3710, 2012.

[33] J. L. Salmeron, "Simulating Synthetic Emotions with Fuzzy Grey Cognitive Maps," Emergent trends in robotics and intelligent systems. Cham: Springer, pp. 39–46, 2015.

[34] K. R. Scherer, Eds, "Facets of emotion: Recent research," in Lawrence Erlbaum Assoiates. New York: Lawrence Erlbaum Associates, 1988.

[35] A. Sloman, "How many separately evolved emotional beasties live within us," in Emotions in Humans and Artifacts. R. Trappl, P. Petta, and S. Payr Eds. Cambridge: The MIT Press, pp. 35–114, 2002.

[36] A. Takanishi, "An anthropomorphic robot head having autonomous facial expression function for natural communication with humans," in Robotics research. London: Springer, pp. 297–303, 2000.

[37] R. E. Thayer, The Biopsychology of Mood and Arousal. New York: Oxford University Press, 1989.

[38] S. S. Tomkins, Affect Imagery Consciousness: Volume I: The Positive Affects. Springer Publishing Company, 1962.

[39] S. S. Tomkins, Affect Imagery Consciousness: Volume II: The Negative Affects. Springer Publishing Company, 1963.

[40] S. S. Tomkins, Affect Imagery Consciousness: Volume III: The Negative Affects : Anger and Fear. Springer Publishing Company, 1991.

[41] S. S. Tomkins, "The quest for primary motives: biography and autobiography of an idea," Journal of personality and social psychology, vol. 41, No. 2, p. 306–329, 1981.

[42] A. M. Turing, "Computing machinery and intelligence," Mind, vol. 59, No. 236, pp. 433–460, 1950.

[43] J. Vallverdú, M. Talanov, S. Distefano, M. Mazzara, A. Tchitchigin, and I. Nurgaliev, "A cognitive architecture for the implementation of emotions in computing systems," Biologically inspired cognitive architectures, vol. 15, pp. 34–40, 2016.

[44] J. D. Velásquez, "When robots weep: emotional memories and decision-making," AAAI-98 Proceedings, pp. 70–75, 1998.

 ISBN: 978-1-61208-839-6