Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 787–793

Emerging technologies for interactive TV

Marek Dąbrowski
Orange Labs Poland
ul. Obrzeżna 7
02-691 Warsaw, Poland
+48 22 699 5706
marek.dabrowski@orange.com

***Abstract*—Advances in web services and open network interfaces enable development of new user-oriented applications in different areas of digital life. Among them, digital entertainment is one of fastest growing networked application domains. Technologies like CDNs (Content Delivery Networks) and HTTP streaming opened the way for new models of TV and video consumption over the Internet. Thanks to emergence of new networking technologies and cooperation paradigms the distinction between traditional TV and the Web is slowly blurring, creating new category of TV entertainment which is more interactive, delivered over multiple screens and inviting users to participate and collaborate.**

This paper presents research on selected social and interactive TV services, focusing on innovations related with incorporation of network interfaces and web protocols. Two use-case scenarios are investigated: one of them explores telco 2.0 interfaces for interactive videos, and the other allows users for social exchanges while watching live sports events. The evaluation results by qualitative user study are presented for the second use-case. A technical architecture is introduced, focused on seamless integration into a typical WebTV platform of telco operator. As important contribution of the paper, some new emerging network protocols and features (like telco 2.0, WebRTC, WebSockets, HTML5 video) are evaluated as potential enabling technologies for future social and interactive TV.

I. Introduction

With the proliferation of IPTV and WebTV content delivery, the TV viewing experience is changing from predominant "lean-back" (passive) viewing habit into the new era of "lean-forward" behavior. People demand more interaction, participation, and social features in addition to just watching the TV. The actors on TV services market (content providers, service aggregators, telco and cable operators) are facing a number of challenges, from the point of view of usage, business and technology, how to integrate the new social and interactive experience into their established service platforms. This paper aims to study selected scenarios for future evolution of social and interactive TV programs, focusing on opportunities brought by new telco network interfaces and web services technologies.

A technical architecture for proposed use-cases is studied from the point of view of telco operator, aiming to integrate new features by leveraging and extending existing TV distribution systems. Special focus is put on OTT (Over-The-Top) content delivery model. The OTT model usually refers to content providers and 3rd party companies who directly access their customers over open Internet. For telecommunication network providers, OTT model starts to be considered as a viable alternative, allowing for addressing additional customers and complementary to the managed IPTV solutions that are still predominant [1]. The OTT content delivery architectures typically use HTTP streaming protocol (e.g. MS Smoothstreaming or Apple HLS) and a CDN (Content Delivery Network) for video distribution over unmanaged Internet with optimal QoE (Quality of Experience).

This paper aims to answer the following research questions:

- From service point of view: what are the realistic and compelling usage scenarios for social and interactive TV?
- From architecture and technology point of view: what are the opportunities brought by new emerging web protocols and network interfaces?

These questions are addressed through analysis of two use-case scenarios, namely:

- Scenario#1: interactive immersive video,
- Scenario#2: social exchanges between spectators of live sports events.

II. Interactive content

The first concept presented in this paper belongs to the category of interactive videos. With interactive video, the movie becomes a kind of immersive game where different story variants may be displayed depending on previous decisions of a viewer. Although interactive videos will probably not enter the mainstream movie industry in the short term at least, they could be appealing for selected types of video production like short artistic movies, instructional scenes, or video games.

The interactive video concept has been known for some years and it has been used for example in DVD interactive games. More recently, it has become quite popular with YouTube service, which offers tools for creating interactive videos. The user may control the video by clicking buttons displayed after the end of particular clip and leading to different variants of the story (see e.g. semi-amateur video [9]). The work presented in article [10] follows the same concept, although removing some of its limitations. The transitions between video clips are seamless and more deeply embedded into narrative of the movie. The authors of [10] provide some guidelines for scriptwriting and shooting interactive videos and identify interaction goals for the viewer: to influence the story, to change the perspective, or to see additional material.

The related works mentioned above have a common limitation: the interaction tool and interface for a user is basically a computer mouse. The viewer has to click a button or an area on the screen to make his decision regarding the storyline. This is not very natural, and feeling of immersion is broken at the time of interaction. On the contrary, interesting video referenced in [11] presents an experiment, where mobile phone has been used as interaction tool. A natural conversation has been imitated between the viewer and the character on screen. The outcome of this conversation has been taken into account in the movie storyline. Interrogated user has been selected among the cinema theater audience. A computer system has called his or her mobile phone and conducted a simple dialog using voice recognition technology.

The approach taken in our project is similar to [11], but not limited to the audience of cinema theater. We propose integration of interactive video system into the WebTV content distribution architecture. Any client located in the Internet could watch and control the interactive video. Like in [11] we use the mobile phone (which is a natural tool for inter-personal communication) to imitate a conversation between the viewer and character on the screen, preserving the feeling of immersion into the fictional story.

A. Usage scenario

Fig 1. exhibits our Interactive Video prototype in use. We have produced a short example movie, not targeted for wide distribution, just for demonstrating the potential of proposed system. The movie protagonist consults the viewer before taking crucial decisions for the storyline. For that, in certain moment the system calls the viewer's mobile phone number, previously entered on a web page. The viewer can hear character's voice on his mobile phone (e.g. a phrase "*tell me if I should take the elevator or use the stairs*"). Then he may answer "*yes*" or "*no*" and in the next video scene the protagonist follows the viewer's advice. All actions are synchronized between the video and real life, i.e. the viewer's phone rings at the exact moment when movie character makes a call; the protagonist starts speaking when the user picks up the call, and during conversation the same voice can be heard on the phone and in the video. The movie character interrogates viewer for his advice in several crucial moments in the story, and the movie has three different endings depending on user's answers.

Thanks to our system the user can feel an illusion of real communication with fictional character while watching the video and is empowered with possibility to influence the story.

B. Architecture and enabling Technologies

A prototype has been developed to demonstrate service concept and serve as playground for evaluation of some new

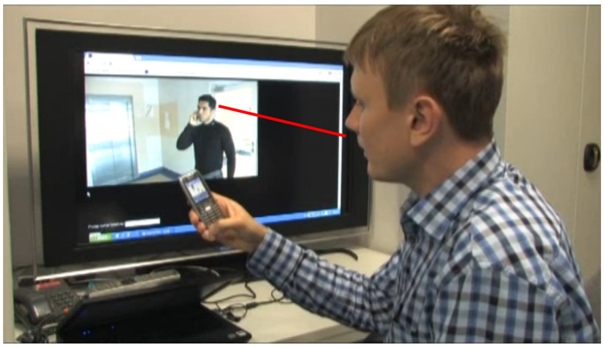

Fig. 1. Interactive video prototype

interesting technologies that may impact future interactive TV. The prototype schema is presented in Fig 2.

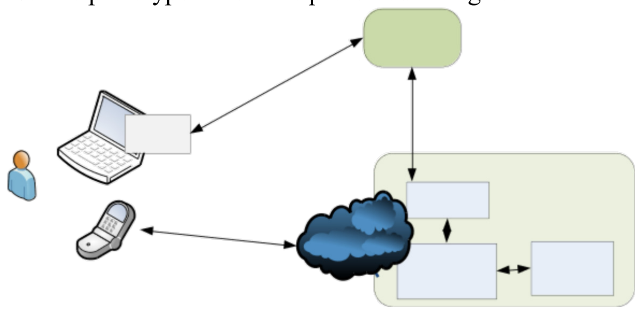

Fig. 2. Architecture of interactive video system

WebSocket and HTML5 video

As design goal, the Interactive Video service should be available for any Internet user equipped with PC computer with a standard web browser, without need of additional software or plugins. Thus, in our architecture the server side hosts logic of interactive movie and controls the sequence of videos that should be played to particular user, depending on his individual interactions with the system. So, the server must be able to send some kind of commands to user's video player (which is embedded in his web browser). This command has to be sent in asynchronous way, i.e. without preceding request from the client. Asynchronous communication is not a usual mode in the client-server world of Internet. There are some techniques to overcome this limitation, e.g. by frequent polling of server by the client, but they can be very inefficient in terms of traffic and processing overhead if one wishes to achieve a timely response. In addition, popular technologies for implementing video player plugins, e.g. Adobe Flash or MS Silverlight, do not allow easily (or not at all) for server-side control of the video player. The player can be controlled from the user-side by mouse clicking on displayed buttons, but the external program has no possibility to change the video that is played, in arbitrary moment.

HTML5 framework [7] comprises some prospecting technologies for solving the limitations mentioned above. The HTML5 video tag enables playing the video natively by a browser. In addition, the video player can be externally controlled by JavaScript code, which in fact facilitates implementing interactive videos over the web. Another HTML5 feature we have investigated is WebSocket [12], an asynchronous protocol allowing the server for sending messages to the client in arbitrary moment. To establish a WebSocket connection, client browser sends a WebSocket handshake request (special HTTP GET message) to the server. Once the WebSocket connection is established, data frames can be sent back and forth between the client and the server in full-duplex mode. Thanks to HTML5 and WebSocket we could implement interactive video scenario as a standard web application, not requiring installing any additional software on the client machine.

For watching an interactive movie, user opens a web page hosted on Interaction Controller server (see Fig. 2). A form for entering a phone number is displayed, as well as "Start" button for launching the video. When a user clicks it, the browser opens WebSocket connection with the Interaction Controller server. A timer process is started for the new user, to control video playback. The timer process will trigger a sequence of user interactions (phone calls at specific times) and video clips, according to pre-programmed scenario.

To play certain video, the Interaction Controller has to send a command to the client browser to open appropriate video file. For that, the server uses WebSocket connection opened at the beginning of the process. The command is interpreted by JavaScript function run by the browser, and it tells browser to stop playing current video and start a new one, available under specified link. The video file is retrieved and played either until it ends, or until next command from Interaction Controller server is received.

Telco2.0 API

The server triggers interaction with a user by invoking advanced Telco 2.0 API functions. Telco 2.0 is a concept of exposing telecommunication services (SMS, call control, location, etc.) through open APIs to external developers. HTTP REST or SOAP is usually used as invocation protocol. Telco 2.0 interfaces are implemented by many network operators and API standardization efforts are recently taken by GSMA organization [13].

Advanced voice API service that we have used allows for establishing a voice call between user's phone and IVR (Interactive Voice Response) machine. IVR is a telecommunication system which controls automatic dialogs with the user. It is typically used by customer relation systems, e.g. when you call a hotline and hear a recorded message *"press 1 to reach customer service, or 2 to report technical problem…"*. In our case, the IVR system has been programmed to play an audio file with recorded voice of the movie character. A user may answer to question asked by the character with almost natural language (in practice, for the purpose of the prototype just simple phrases were used like *"yes/no"*). The ASR (Automatic Speech Recognition) system is then engaged for voice analysis and matching the answer with pre-configured phrases. Finally, the Interaction Controller server is notified about the user's answer to the interaction question. The server, knowing the user's response, chooses appropriate variant of the story and accordingly commands playback of a video file using HTML5 features as described previously. The interaction process is executed in real-time, i.e. there is no perceived delay between the decision expressed in phone conversation, and effect observed in the video.

Evaluation

The concept has not been tested by real users, as we recognized that user's perception and perceived usage benefit depends a lot on a story that is being told by the interactive video. As we only had possibility to produce a simple amateur video as a proof of concept, the user tests would be too much biased by assessing the video (narrative and artistic quality) instead of potential offered by new technology. However, internal tests done with the prototype gave us some interesting insights on the technical and usage aspects.

On the technical side, new web-oriented technologies and protocols like HTML5, WebSockets, Telco2.0 API, have proved their merits for supporting interactive video scenarios. The proposed server-based architecture correctly manages execution of interactive video and allows for providing it as a service to WebTV users. The voice API has been confirmed as powerful tool for carrying out automated voice-based dialogs with users.

Some usage tests have been performed, however only by project members and thus the trustfulness of results is limited. Anyway, based on the feedback of this limited experiments we can confirm that the interactivity is fluid and switching between consecutive videos is almost seamless. Feeling of immersion in the story is genuine and experiment participants could admit that conversation with fictional character was really funny and enjoyable for them. As a drawback we can mention that the automated dialog and voice recognition technology is still not perfect when user says a phrase or sentence that is not understandable or unexpected. As a hint for future work, such situation should be covered in the interactive video scenario, e.g. the movie character should keep on asking viewer for repeating his answer in the case it had not been properly recognized by the system.

III. Social Interactions Around Sports Events

Social TV is recently an often discussed trend in media industry. In fact, this term covers a wide range of scenarios and applications. As a general consensus, Social TV refers to services where TV watching experience is augmented with some forms of inter-personal and inter-audience communication.

The scenarios where content (i.e. a TV show or a movie) is a trigger for social communication, have been studied in numerous related works. For example, paper [2] distinguishes two types of communication around content: synchronous (while watching TV, e.g. through text or voice chat) and asynchronous (after watching, e.g. by leaving notes or recommendations for people who will watch the same content at another time). The authors of [2] have analyzed impact of TV genres (like movies, news, sports, soap operas) on social activities of people. Apparently, sports is one of the genres that is the most "talked about" while watching, i.e. is a good candidate for social TV system supporting communication in synchronous mode. Intuitively, we can expect that watching sports triggers intensive emotions, and it is natural for us to share these emotions with others, especially friends and relatives. In addition, contrary to films or documentaries, sports do not have strict plot-oriented structure and do not require full viewer's attention all the time. There are usually some breaks and low-activity periods when the action on screen is not so absorbing and the social exchanges at that time would not interrupt watching experience so much.

The social communication aspect can be realized through (1) existing social networking services like Facebook or Twitter, (2) a dedicated application, or (3) enhanced TV service delivery platform. As an example of the first approach, paper [3] presents exhaustive report on how Twitter has been used by producers of popular TV reality show X-Factor for enabling inter-audience communication during and after the show. Interaction with the audience by interactive communication channels and 2nd screen has become an intrinsic part of TV show format. The example of second approach is GetGlue [4], a popular application for sharing and exchanging comments while watching TV shows.

In our approach, the communication features are even more closely integrated into the TV viewing experience thanks to implementing them as part of live TV content distribution architecture. We propose a Social TV system that is specially designed for sports programs and adapted for using during live TV transmissions. Our solution enables social exchanges among friends and fan communities.

A. Usage scenario

In our scenario, social communication around sports events has two "dimensions". The first one concerns exchange of messages, photos and user-generated videos between people who are physically present on the tribunes (e.g. of a football stadium), and those who stay at home and watch live transmission on their TV set or PC computer. Using special mobile application, spectators at the stadium may share emotions and comments with viewers at home. After launching the application, user may select the "event" he attends to. The "event" is configured by service provider and it corresponds to particular match or concert. Then, the application role is twofold. First, it allows for browsing and watching the messages, photos and videos sent by other spectators of the same event. Each message may be commented (by adding text) and ranked by clicking on an icon of a thumb. Second, most important, the application allows for quick upload of your own content after clinking on "send message", "send photo" , or "send video" icons.

Uploaded messages are displayed in real-time on top of the video watched by users of WebTV service. Special video player has been developed using MS Silverlight framework. It displays incoming messages as a scrolled list on the right side of the screen, with preview of picture or video content appended to the message. Viewer may enlarge each message to see its rank and associated comments. He may also add his comment and rank the message in the same way as users of mobile application. Thanks to our system people at home may have a glimpse of stadium atmosphere, which is shaped by fans supporting their players. Example screenshot of WebTV player, displaying a gallery of photos provided by fans at the stadium, is presented in Fig 3.

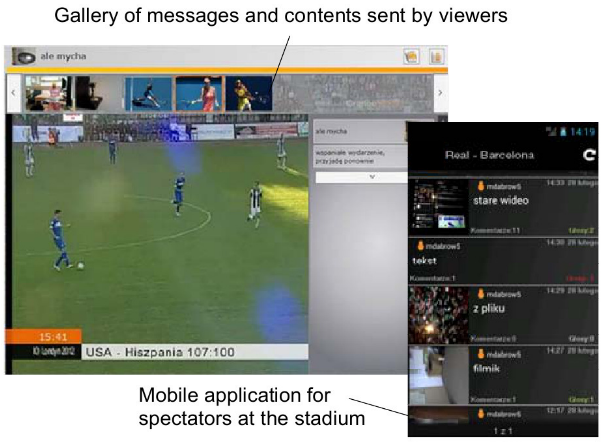

Fig. 3. Sharing photos/videos between spectators at the stadium and viewers at home

Remark that our service concept is a bit similar to FanFeeds, described in article [5]. FanFeeds is a system for aggregation and presentation of additional content-related media on 2[nd] screen while watching the program. This supplemental information can be contributed by people (from our social group, or strangers) who are watching the same content. The authors of [5] present the prototype concept and results of field trial, which thoroughly investigated people's motivations and incentives to communicate over content. The need to exchange comments during watching the content has been confirmed through trial results. Basically, we enable sharing comments between viewers of a live event in similar way as FanFeeds. However, in our approach we specifically focus on connecting two groups of people: spectators physically present at the stadium, and TV viewers at home. In this way, a fan community can be created by spectators of a sports event, regardless if they attend it physically or at home on TV.

Another dimension of social exchanges around sports events, that we consider in our system, is real-time interpersonal communication. With our service, friends or relatives can watch the televised event together despite being apart in their homes. Real-time communication by text chatting while watching TV has been often mentioned in literature and for example article [6] presents an advanced social TV system based on chat. Real-time voice and video communication is less popular, as it is more challenging from the usability and also technical point of view (a user device equipped with camera is certainly required). The experience of videoconference while watching TV may be replicated by using external application, like e.g. Skype, launched on a PC computer. We take a different approach, by integrating videoconference into the TV content delivery system, providing user with a single interface to access TV service together with the social features. PC and camera is currently required to run the service, while in the future an STB equipped with camera could also be envisioned. Thanks to group management and presence service a user may see that his friend is watching the same sports program, and invite him to watch together. Fig. 4 presents exemplary screen layout with two participants watching together a tennis match.

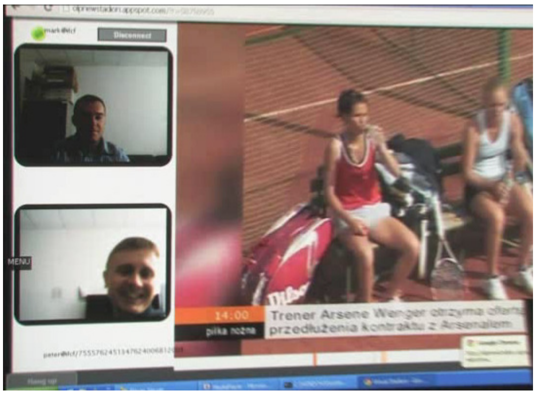

Fig. 4. User-to-user communication while watching TV show

We can say that people who have joined the service are gathered on a "virtual stadium" and constitute a kind of social circle of virtual supporters. They may communicate by voice an they can see each other. In addition, they may take part in quizzes or small games. In our developed prototype, the quiz questions are moderated by service operator and sent to all participants at the same time.

B. Architecture and enabling Technologies

A proof-of-concept prototype has been developed to verify viability of the use-case (concerning both dimensions of communication as described above) and maturity of enabling technologies. It extends typical OTT-based content delivery system with components for managing social exchanges and inter-personal communication. The prototype schema is presented in Fig. 5. On the server-side we distinguish the WebTV platform, which groups content preparation and distribution functions (encoding using H.264 standard, and streaming with adaptive protocol MS Smooth-streaming). The streams are delivered over CDN (Content Delivery Network). The social TV features are provided by two servers. The "Virtual Stadium" server provides clients with a main service web page and manages real-time communication: users identification and presence, as well as groups and videoconference participants matching. The videoconference media is exchanged in peer-to-peer mode between communicating clients by WebRTC technology. The "UGC server" handles functions related with non-real time communication mode. It aggregates messages submitted by user's smartphones (an Android application) and stores associated contents as JPG and H.264/MP4 files. For enabling unified display of UGC messages overlaid on TV content, a special video player has been developed with Microsoft Silverlight.

Although most of the technologies used for the prototype are standard and widely known, the WebRTC merits some attention as it is relatively new and can be considered as very promising for communication services over the Internet.

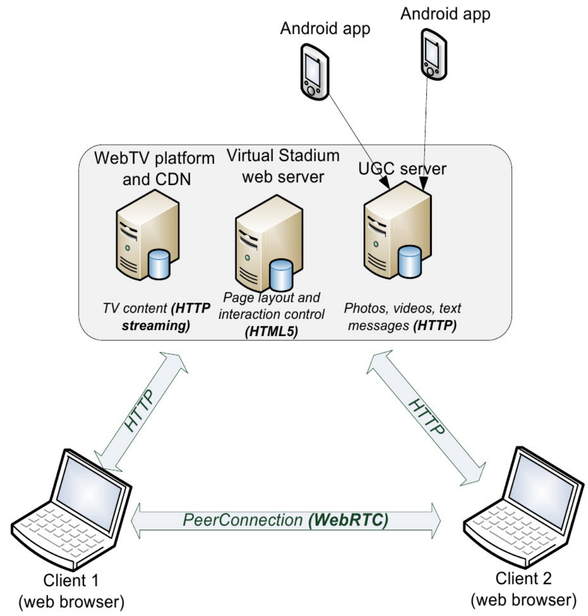

Fig. 5. SocialTV prototype architecture

WebRTC

WebRTC is part of HTML5 framework [7] and it enables embedding real-time communication component within a web page. The communication session is handled directly by web browser, without need of any software plugin. WebRTC [8] defines a standardized browser API that can be invoked using JavaScript by web applications for services such as voice and video calls. WebRTC is subject of ongoing standardization works in IETF and W3C.

The client (see Fig. 5.) opens a "Virtual Stadium" web page on his PC computer. A Silverlight player downloaded from the web page provides access to live TV content through connection with WebTV platform and CDN (Content Delivery Network). When a user decides to connect with his friend, Virtual Stadium server is invoked to generate a session token that is shared by web browsers of communicating parties. A WebRTC peer-to-peer connection is then established between them, for direct exchange of videoconference media (audio and video), all displayed on the same web page as TV stream.

Thanks to WebRTC technology the videoconference service could be seamlessly and easily integrated into the WebTV content distribution architecture.

C. Evaluation

Apart from assessing technical maturity, important goal of research work was to evaluate potential market adoption of the Social TV service. Qualitative tests with real users were conducted for this purpose, using methodology of Focus Groups Interviews (FGI). Four sessions were carried out, each with 6 participants recruited by external agency with regard to typical Social TV user profile that is: age 21-40, people watching matches, concerts, etc. either live or via web TV, active users of social networking, video-chat applications, internet chat rooms and discussion forums. The goal of the test was to survey reaction of potential users to the Social TV service concept, determine factors which encourage people to use the services and identify any usage barriers. During interviews, service concept and demo videos were shown to participants, followed by a guided discussion on perceived benefits and flaws of the presented concept.

We can conclude from the interviews that users positively perceive the service as meeting their need of sharing comments and emotions while watching live events on TV. The opinions gathered during experiments confirm findings of [2], saying that sport is one of the TV genres that especially trigger social exchanges. Watching live sports is related with strong emotions, which are best experienced in a group. We may quote of the test participants: "*When you're alone, you don't feel the thrill. You have no one to experience it with.*".

Referring to possibility of sharing messages and photos between spectators at the stadium and viewers at home, interviewed persons appreciated the feeling of being a part of live event, thanks to access to additional content sent by participants in real time. Moreover, they liked the possibility of seeing more than is shown by the camera in a typical broadcast: "*At such an event, you cannot be everywhere at once. And we all know that everybody pays attention to slightly different things.*"

The aspect of sharing emotions and experiences has been appreciated by test participants. The impression of watching the event in the company of others, being in connection with the fan community, is important.

The respondents pointed also to some limitations of the concept. A bias has been detected between the benefit perceived by viewer at home, who enjoys additional supplemental content, and benefit of people at the stadium, who may not have sufficient motivation to use our service. We think that possible ways to increase the perceived benefit for users at the stadium are: a competition for spectators (win a prize for best photo from the tribunes), or adding a gamification scheme (collect points and rewards for sharing). Test participants advocated also for possibility of creating user groups and sharing comments within a closed social circle instead of general population of TV viewers. Comments collected during the tests certainly give us interesting hints for future development of the service. A field trial is planned in spring 2013, where the prototype will be evaluated during a real football match event.

As main benefit of Virtual Stadium use-case people mentioned possibility to share opinions and comments with friends immediately during live TV broadcast. The interviewees emphasized the difference between exchanging comments on the event while it was still on, with sharing them after the event is over. The respondents agreed that they saw a greater need to share emotions during the event than after it was over.

Although some concern has been raised that interactivity may disturb perception of the event, participants noticed that while watching sports there are usually some less exciting moments, when conversation with friends would be hardly disturbing for the viewing experience.

Some test participants have previously used other tools for communicating with friends while watching TV: ("*I once watched a match over Skype with a friend who was in South Africa then.*"). Nevertheless, the idea of all-in-one service, integrating several currently used tools has been appreciated as making the experience easier to use. The respondents liked also the fact that it would be available simply in a web browser, without need to install additional software.

Quizzes and games during the match have been judged rather controversial. People find it annoying if the quiz question pops up suddenly on the screen, covering the broadcast. The quiz component has to be designed with great care, being less intrusive and providing sufficient benefit for the users to participate (rewards, competition, possibility to show-off your knowledge in the area, etc.).

Another critical comment concerned current limitation of videoconference to only two participants. Sometimes there is a need to comment or discuss in a larger group. In future works the videoconferencing system should be extended for multiple users.

IV. Conclusion

The paper has discussed several exemplary realistic scenarios for TV services of the future, focusing on opportunities brought by new advances in network interfaces and web protocols. The interactive video use-case assumes imitating a voice dialog between the viewer and fictional character of the movie. Social TV for sport programs allows for two types of inter-audience communication: sharing of comments and photos between spectators at the stadium and viewers at home; and videoconference connection between friends while watching live sports event. The scenario has been evaluated, with positive feedback, by small scale user test.

For all presented use-cases a technical architecture has been designed and prototyped, assuming full integration with OTT-type content delivery architecture of a telco operator. The service logic is implemented by dedicated server components, and standard web browser is used as unified client device for content consumption as well as for social features. In the case of interactive video, mobile phone is employed as second device, using its most natural function, that is a voice call.

Several emerging technologies have been evaluated considering their potential for enabling interactive TV. Among them, HTML5 and WebSocket overcome limitations of traditional web architecture for interactive videos controlled from the server side. Advanced Telco 2.0 interfaces can be used to engage with the viewer through his phone. Mobile applications and open web services interfaces allow for quick and easy upload of user-generated text and contents. Finally, WebRTC is a promising framework for integration of content delivery and inter-personal communication within a web page.

REFERENCES

[1] VideoNet report: The hybrid network operator, March 2013
[2] D. Geerts, P. Cesar, D. Bulterman, The Implications of Program Genres for the Design of Social Television Systems, uxTV 2008, USA
[3] M. Lochrie, P. Coulton, Sharing the viewing experience through Second Screens, EuroITV2012, Berlin, 2012
[4] GetGlue application: http://getglue.com/ (last accessed 10.07.2013)
[5] S. Basapur et al., FANFEEDS: Evaluation of Socially Generated Information Feed on Second Screen as a TV Show Companion, EuroITV2012, Berlin, 2012
[6] F. Martins et al., SentiTVchat: Sensing the Mood of Social-TV Viewers, EuroITV2012, Berlin, 2012
[7] W3C HTML5 specification, http://www.w3.org/TR/html5/
[8] WebRTC initiative web page: http://www.webrtc.org/ , Decemeber 2012
[9] Interactive horror, YouTube video: http://www.youtube.com/watch?v=D2mMIxXGvEE (last accessed on 10.07.2013)
[10] Interactive Movie Demonstration Die Hobrechts, Drei Regeln (Three Rules). In Adjunct proceedings of EuroITV2013. Video available at: http://preview.3regeln.de/player_en.html (last accessed 10.07.2013)
[11] Interactive Horror Movie Last Call, YouTube video: http://www.youtube.com/watch?v=386VGKucWDo (last accessed 10.07.2013)
[12] IETF RFC6455: The WebSocket Protocol, December 2011
[13] http://www.gsma.com/oneapi/