Coping with the Challenges of Delivering Multiple Sensorial Media

Coping with the Challenges of Delivering Multiple Sensorial Media

Estêvão Bissoli Saleme
Federal University of Espírito
Santo, Brazil

Dr. Celso A. S. Santos
Federal University of Espírito
Santo, Brazil

Dr. Gheorghita Ghinea,
Brunel University London,
UK

Multiple Sensorial Media (MulSeMedia) systems transcend the traditional senses of sight and hearing, adding smell, touch, and taste into multimedia applications to create more immersive experiences for the users. Here, we provide a picture of the challenges and requirements for mulsemedia delivery and present a solution to cope with software and hardware heterogeneity.

Imagine a chemistry experiment being streamed to thousands of kids on social media. Imagine the kids wearing VR headsets, all focused on the experiment about to begin. A scientist in a white coat enters the room and selects two receptacles from the shelves containing chemical goodies, one containing a green liquid, the other a yellowish substance. The scientist takes a pipette, measures the right volume of the green liquid which she then deposits at the bottom of a glass tube. She then carefully cuts a piece out of the yellow substance, takes hold of it using a pair of tweezers, and then drops it in the glass tube. FLASH! SMOKE! SMELL! The kids gasp in awe – for not only did they *see* and *hear* the exothermic (heat emitting) chemical reaction of the experiment in the VR headset, but they also *smelt* the pungent odor emitted as a result, felt a *whiff* of hot air blowing through their hair, as well as a slight *vibration* on their bodies. Fantasy?

Perhaps not, for the students have just experienced an immersive *mulsemedia* experience. Almost 20 years ago, Purchase[1] suggested a model with three dimensions to deal with the definition of multimedia communication supported by semiotic concepts. The first and second dimensions were concerned with the nature of the semiotic sign and the organization (syntax) of them. The third was dedicated to the senses used for the multimedia communication. Back then, the third dimension was modeled by only taking into account two values, visual and aural. Still in the 90s, Arens, Hovy, and Vossers[2] also considered aural and visual as the sole values of media types for multimedia presentations. VR-based mulsemedia - multiple sensorial media - unlike multimedia content, involves novel, non-traditional media targeting one or more of the sense of

smell (olfaction), touch (haptics) and taste. Whilst a few decades back having such digital mulse-media applications might indeed have been the realm of fantasy, recently there has been a prolif-eration of commercial, off-the-shelf devices which can now produce such novel user sensations.

Coming back to our chemistry scenario above, there is no reason why the VR headset worn by the children cannot contain a small tube mounted where it rests on the bridge on the nose which can dispense directed, synthesized aromas to the user, such as the prototypes of Nakamoto et al.[3] and Covington, Agbroko, and Tiele[4]; one could just as easily imagine a small ventilator fan inte-grated with the headset near the user's temple which takes advantage of the heat generated by the device to blow shots of warm air, whilst the vibration felt by participants is that produced by a haptic vest worn by them and which vibrates in sync with audio impulses received. To make matters even more interesting, 360º videos can also be used to enhance the user's sense of im-mersion, whilst the glasses themselves can contain eye trackers which follow the user's gaze, allowing them to focus and zoom in on regions of particular interest through gaze-contingent dis-plays.

Mulsemedia applications, though, have characteristics which distinguish them from their multi-media counterparts. Whilst the latter predominantly contain media which are either continuous in time (such as video or audio) or space (such as text), the concept of continuity is rather hard to pinpoint for mulsemedia applications. Both olfaction and taste are characterized by *lingering* - the tendency of aromas to linger in the atmosphere and of tastes to linger on the tongue mean that traditional spatiotemporal continuity makes little sense here. When *integrating* such linger-ing stimuli with traditional audio-visual ones, in a production-distribution-rendering context, challenges multiply.

In this article, we describe fundamental challenges for delivering multisensory effects to hetero-geneous multimedia applications and hardware taking into account the sometimes-quirky aspects of mulsemedia. To do this end, we will be using the PlaySEM platform, which is capable of de-livering mulsemedia content to heterogeneous applications. The platform was created by Saleme and Santos[5] in order to initially deliver sensory effects such as wind, lighting, and vibration, specified in MPEG-V, to multimedia applications by reusing the PlaySEM Sensory Effects Ren-derer (SER), the cornerstone component of the PlaySEM architecture. PlaySEM[6,7,8] has evolved to support different profiles of integration following architectural and design patterns to accom-modate a multitude of communication and connectivity protocols and sensory effects metadata standard according to the needs of multimedia applications and devices available in the user's environment. We opt for PlaySEM because it has been favorably applied to diversified scenarios.

We claim that multimedia application developers should not lose focus on their core business whilst adding support to multisensory effects in their applications based mainly on the principles of reuse.

CHALLENGES AND REQUIREMENTS FOR MULSEMEDIA DELIVERY

Portraying – and experiencing – mulsemedia content has to accommodate a wide spectrum of situations, conditions, and constraints. Recent work has identified major obstacles and issues which the deployment of mulsemedia systems have to overcome. They include integration, syn-chronization, control of intensities, the use of wearable, portable and other heterogeneous de-vices for delivering sensory effects, remote delivery, and rendering of mulsemedia contents[9,10]. Additionally, mulsemedia systems have to be able to adapt to incorporate different operating sys-tems, presentation interfaces, standards, connectivity and communication with applications and devices. Due to the continuous evolution of technology, reusability is key to coping with hetero-geneous settings. Figure 1 presents a picture of the challenges identified and how the require-ments are spread between them. This point of view provides mulsemedia developers with a summary of what they have to cope with.
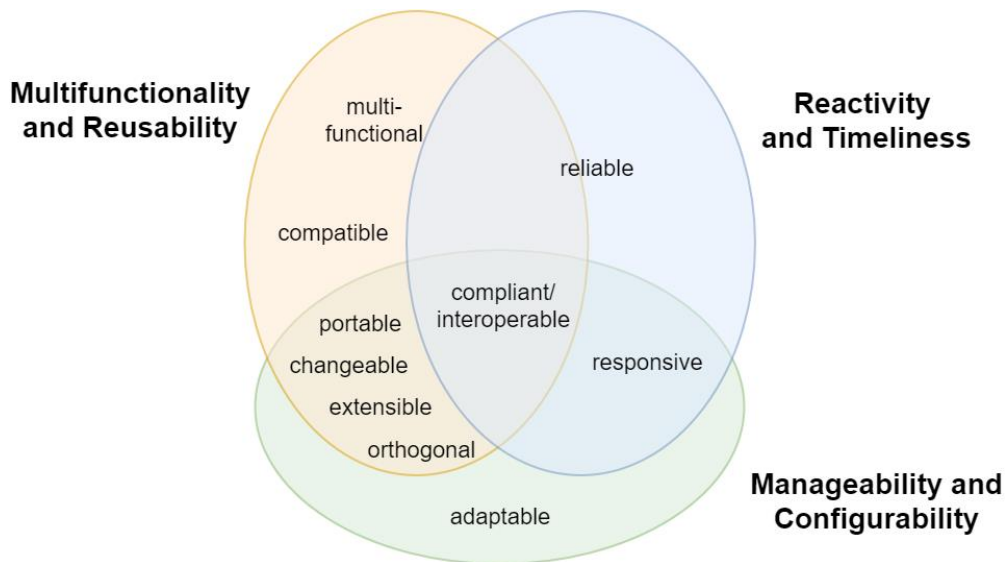
Figure 1: Relationship between challenges and requirements for delivering sensory effects to heterogeneous multimedia applications.

After assembling a mulsemedia solution for a client, a shrewd developer might want to reuse it for other projects. Thus, it must be developed targeting the challenges of *multifunctionally and reusability*. *Multifunctional* means that the mulsemedia system will provide functionalities for multimedia applications to add multisensory effects into it, being able to work with timeline (e.g. media players) and event-based (e.g. games, AR/VR software, and interactive applications) multimedia applications. Furthermore, mulsemedia solutions need to be *compatible* with their multimedia counterpart, providing reusable implementations of services to communicate with it. *Portability*, the need to run mulsemedia applications on a wide range of platforms, carries with it a requirement for *compliance* with standards or rules enabling interoperability. It is our view that architectural patterns are the main enabler of reusability in mulsemedia systems. By adopting a layered architecture, multimedia applications could use common services from an interoperable mulsemedia system that decouples presentation from processing and rendering, for instance. Such an approach would go somewhat towards catering for the distinct requirements of spatiotemporal media and those of lingering media in an integrated fashion.

Another challenge is related to *reactivity and timeliness*. Extending our chemistry example above with mulsemedia interactive scenarios, in which students deepen their knowledge of chemistry– and having fun in the process – by dabbling with various chemical substances in a VR/AR context, the fact that mulsemedia systems should be able of handling real-time requirements is unsurprising. For any application to be worth its salt, the ability to consistently produce the same results under temporal constraints – *reliability* is also necessary, whilst *responsiveness* mandates that delays should be minimized so as multisensory stimuli are rendered and engage the user within a user-acceptable time window. Indeed, synchronization does not have to be 100% precisely accurate but any imperfection should not be noticeable by humans[11].

Coping with heterogeneity of devices, running a variety of operating systems and being connected via a plethora of communication protocols represents a substantial challenge for mulsemedia systems. Moreover, specific types of multimedia applications might require the use of different communication protocols on the same mulsemedia application - for instance, a multimedia web video player could use the WebSocket protocol to communicate with mulsemedia renderers whereas a desktop video player could use UPnP. In this fashion, low cohesion and high coupling may entail a deep change in the code. Mulsemedia systems should thus be able to work with heterogeneous multimedia applications without difficulty, hence dealing with *manageability and configurability*.

We also posit that mulsemedia systems should be *changeable*, *extensible*, *orthogonal* and *adaptable,* and architectural and design patterns can again be used to this end. The ability to be

changeable features mutability in their structure without interfering with the adjacent layers. Extensibility plays an important role in changeability in the sense that the design of the system takes future growth into consideration. Orthogonality at the architectural level will also aid the developer to anticipate possibilities to combine different components. Adaptability makes available mechanisms to customize the system to their needs without needing to change the system, that is, mulsemedia systems should be designed so as developers can configure and extend their applications according to their specific needs.

ADDRESSING THE CHALLENGES WITH THE USE OF THE PLAYSEM SER

To cope with the aforementioned challenges and meet those requirements we explore the PlaySEM SER. It is the central component of the PlaySEM platform, a decoupled open source software written in Java to work with multisensory applications[5,6,7,8]. It deals with heterogeneous applications (i.e. timeline and event-based multimedia applications) abstracting the processing of sensory effect metadata sent by them, making a bridge between applications and hardware. A clear benefit of this approach is that existing multimedia applications do not have to drive heterogeneous hardware to deliver sensory effects, once it is a duty of the PlaySEM SER. To enable it, a layered architecture is provided based on the lessons learned from previous experiences. This architectural arrangement supports the addition of new technologies without changing adjacent layers. By stowing the internal layers from everything except the adjacent outer layer, layered systems reduce coupling over various layers along these lines pushing ahead evolvability and reusability[12]. Figure 2 provides a holistic view of its architecture through a layered division.
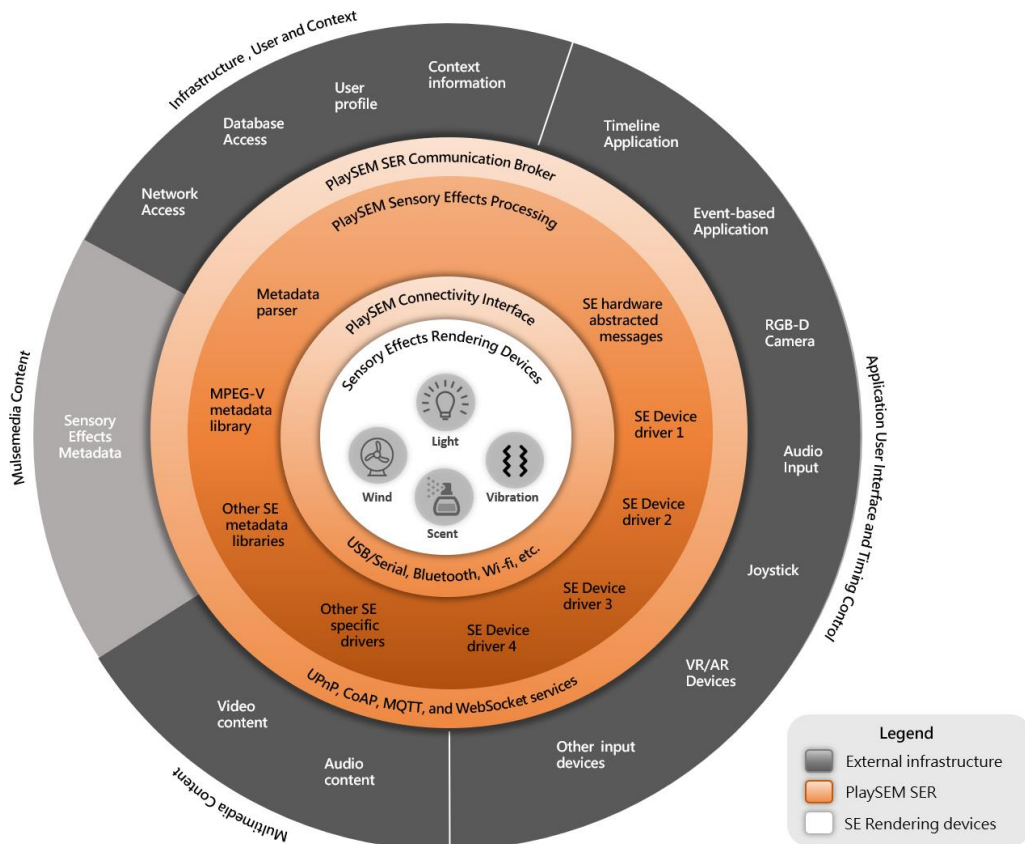

Figure 2: Holistic view of the PlaySEM SER's tiers through a layered architecture.

The outermost layer represents multimedia applications (timeline or event-based), their devices (AR/RV, sensors, etc.), construction, and associated content. The external infrastructure is isolated from the PlaySEM SER to separate concerns. Thereby, the latter can focus exclusively on mulsemedia keeping apart issues such as adaptive streaming, buffering, video encoding and decoding, user interface, among others. Recurring to our chemistry experience, the VR headset would fit here. The PlaySEM SER then caters equally for these multimedia applications to add support to sensory effects regardless of the type of the application. The goal is to account for *multifunctionality*. Though this generalization and decoupling could suggest timing issues such as the lack of synchronization for timeline applications and delay for event-based ones, the studies of Saleme and Santos[5] and Saleme, Santos, and Ghinea[8] brought evidence that it is negligible in this fashion. Furthermore, this layered architecture could be seen as a Mulsemedia Delivery Architecture (MDA) by just swapping the PlaySEM SER for an equivalent software.

Common services for processing metadata and controlling the delivery of the sensory effects to the devices are provided by the neighbor layer called PlaySEM Communication Broker, in sync with the multimedia application. Taking into account that multimedia applications can run in different environments, this layer provides these common services through different communication protocols such as UPnP, CoAP, MQTT, and WebSocket to make it *compatible* with a plethora of multimedia applications. Therefore, the VR headset could communicate with the broker through one of these protocols to convey sensory effects about the chemistry experience to the PlaySEM SER. These protocols also offer mechanisms to ensure that the messages are correctly delivered to the PlaySEM SER, thus providing *reliability* to produce the desired results. Moreover, considering that some of them are lighter, it makes the results more *responsive* to minimize delay.

The PlaySEM Sensory Effects Processing layer deals with metadata processing converting them into sensory effects hardware abstracted messages to be, in turn, translated into specific commands for controlling the devices at runtime. The requirement for *compliance/interoperability* calls for multiple support to sensory effects metadata. Currently, the MPEG-V standard is widely known, hence the PlaySEM SER supports it for light, vibration, wind, and scent. However, it could be replaced by other efforts in the future, either for avoiding the payment of royalties or increasing expressiveness of sensory effects metadata to include, for instance, support to taste description. The PlaySEM SER does not hinder this replacement by offering a mechanism to define a strategy to process sensory effects metadata before running the system. It also enables *portability* at sensory effects metadata level and leads to meet the requirement for *changeability*, in this case, to change the system without interfering with other layers.

Between the Sensory Effects Rendering Devices and the PlaySEM Sensory Effects Processing layer, there is the PlaySEM Connectivity Interface. This layer makes available for use a set of connectivity protocols so as the rendering devices can be connected through different protocols. For example, instead of a wired connectivity one may want to use a wireless connectivity when, for instance, wires are unwieldy like in the case of the haptic vest worn by the kids in the chemistry experience to produce vibration. Therefore, a connection with a device can be established according to the need of the multimedia application. This takes into account the requirement for *orthogonality*, allowing developers to anticipate possibilities to combine different components.

The innermost layer is the Sensory Effects Rendering Devices where the actuators are grouped. The synthesized aromas, heat, and vibration effects that the kids would be immersed in the chemistry experiment would come from the devices placed here. For each device, the PlaySEM SER provides a delay parameter that can be individually set up in order to compensate for delay inherently generated by slower devices. For instance, a heating device that takes 5s to deliver a hot air can be started 5s before it is described within the sensory effects metadata. It might be useful to adjust different synchronization levels without changing the sensory effects metadata or even the application.

In many cases, a developer is able to change the connectivity of a device, swap the sensory effects metadata standard, use different communication protocols with multimedia applications, and replace a device without changing the code. When it is not possible, the PlaySEM SER assumes the shape of a framework to make the solution *extensible*. To add new services, other communication protocols, other metadata standards, other connectivity protocols, and new devices, the system makes use of the patterns abstract factory, singleton, facade, observer, strategy,

and template method. They help to avoid many problems such as creating an object by specifying a class explicitly, dependence on hardware and software platform, dependence on object representations or implementations, tight coupling, and extending functionality by subclassing. The patterns are combined with a configuration mechanism accounting for *adaptability*.

In layman's terms, developers of the chemistry experience could focus on the VR experience whilst adding support to sensory effects through the PlaySEM SER by adjusting it to its profile.

CASE STUDIES

To demonstrate the applicability of the PlaySEM SER, we provide four case studies in which different settings were used to deploy a mulsemedia solution. The first one presents a video player working with the PlaySEM SER. The second one describes the integration of an interactive application controlled by gestures and the PlaySEM SER. Then, the third brings a case study where the temporal aspect of that integration was investigated and used to improve response time in networked event-based mulsemedia applications. Finally, the fourth one depicts a scenario where the PlaySEM SER was employed in an IoT-based architecture for mulsemedia delivery, adapted with different devices and exposed to user assessment. Though some of these case studies were experiments carried on previous versions of the PlaySEM SER, the current solution keeps the same functionalities and additionally takes new requirements into account.

Movies Running on the SE Video Player

The use of sensory effects along with movies is not a novelty. Nevertheless, the work of Saleme and Santos[5] brought an effort to decouple multimedia applications from mulsemedia systems. It presents PlaySEM, a platform that established a new approach for simulating and rendering sensory effects and operates independently of any Media Player, and that is compatible with the MPEG-V standard while taking into account the requirement for reuse. They realized that the use of standards to integrate the real world with the virtual world and vice-versa was crucial, coming up with the MPEG-V standard as a bridge to integrate the systems over a network and using the UPnP protocol. Figure 3 portrays the user's environment playing a movie on the Sensory Effects (SE) Video Player integrated with the PlaySEM SER running on a mini-pc, which in turn, is connected to a microcontroller Arduino that handles a LED strip to produce lighting effects, PC Fans to create wind, and vibration motors to generate haptic effects.
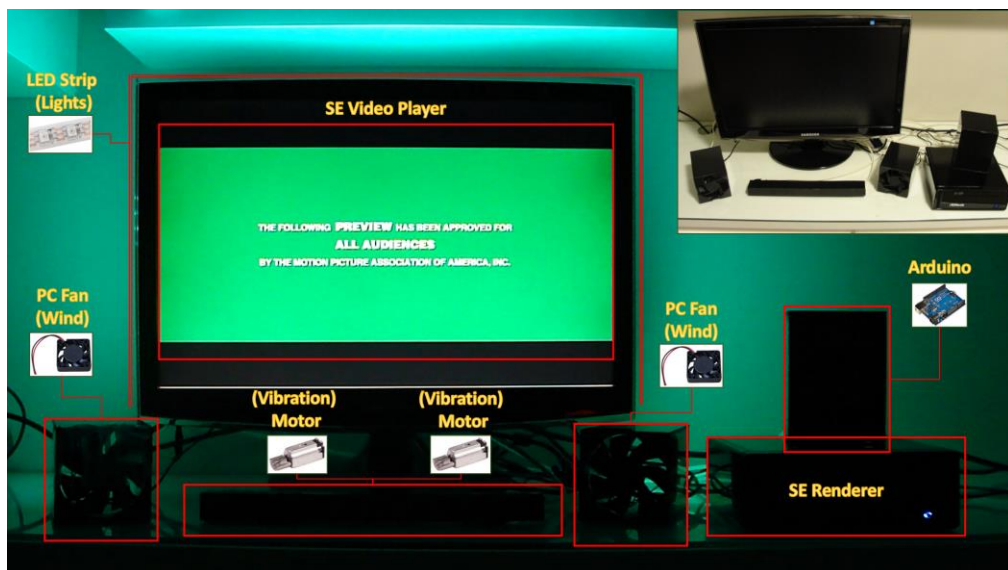

Figure 3: Environment showing a movie on the Sensory Effects Video Player integrated to the PlaySEM SER and its connected devices.

The evaluation presented in their work consisted of experiments on the decoupled approach, for verifying if it affected the processing of messages sent by the multimedia application and whether the proposed architecture had a relevant impact on the loss of synchronization between the clocks. They found that the decoupled use of the renderer did not cause any relevant communication delays, with values below 5ms. Moreover, it was shown that the PlaySEM SER time had a negligible impact on synchronization loss, with maximum losses of 34ms.

Integration of a Gestural Interactive Application

Santos, Neto, and Saleme[6] proposed an approach to integrating event-based multimedia application controlled by gestures with an interactive environment with sensory effects. They argued that the development of an interactive environment is demanding in the sense that it takes time and effort for coping with different software components and information related to sensors and actuators. Three main tasks were defined to accomplish it. First, the data needed to be captured through sensors. Second, there was a processing stage in which the application had to recognize gestures and then triggered desired actions. Subsequently, these actions had to be performed by sensory effects actuators. Along the work, the authors also proposed a method to achieve it based on the concepts of the behaviorism, mainly on the principle of stimulus-response. After recognizing a gesture (stimulus), a set of sensory effects occurs (response). The sensory effects were described through the MPEG-V standard and were triggered whenever a stimulus was associated with it.
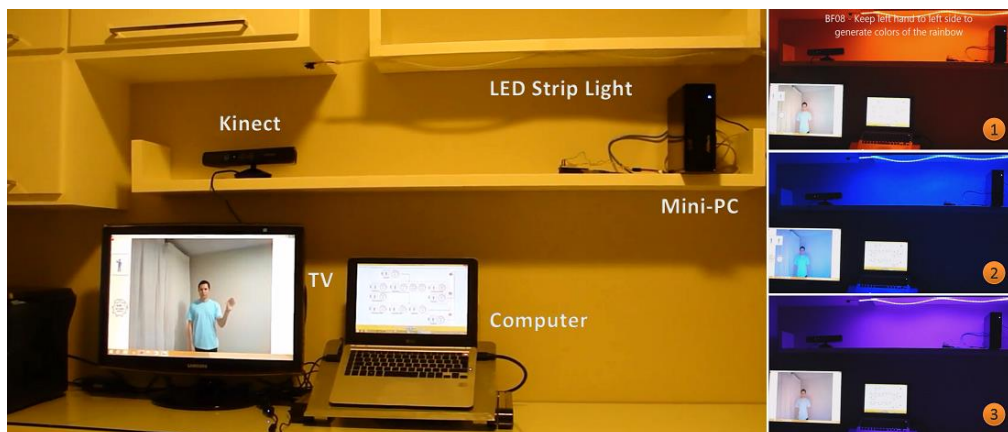

Figure 4: Experimental environment and its components for integrating multisensory effects into interactive environments.

The method to validate the proposed approach was a case study, which integrated multiples sensory effects to an interactive environment based on gesture recognizing. Figure 4 shows a Kinect to capture the gestures, a computer to recognize them, a TV to show the user a feedback of the interaction, a mini-pc running the PlaySEM SER, and a LED strip to produce the lighting effects. The steps one to three on the right side of the same figure present the outcome of the experience after recognizing a gesture, triggering actions associated with MPEG-V scripts, and finally rendering the sensory effects. The environment was described through a set of stimulus, actions, and behaviors following their approach. The case study intended to simulate a theatrical play controlled by gestures and revealed the possibility to develop and integrate interactive environments with mulsemedia applications, providing a more natural and enjoyable experience for the users.

Temporal Aspects for Mulsemedia Event-based Applications

Identifying stringent numbers for tolerable delays is quite strenuous to do and rather dependent on the context of the interaction. Many studies have identified limits in varied setups with dis-

tinct types of sensory effects. Nakamura and Miyashita[13] found out that electric taste can be presented with visual stimuli with a difference between [13ms, 837ms]. Murray et al.[14] discovered different thresholds depending on the scent type, that is, foul scent [0s, +15s], spicy, fruity, flowery scents [-10s, 10s], and burning scents [-5s, +10s]. For haptic accompanied by videos, Yuan et al.[15] indicated [0s, 1s] as a tolerable range. On the other hand, when haptic is presented with a head-mounted display, like the work of Adelstein, Lee, and Ellis[16], system latency must be no higher than 17ms. As pointed out by Kim, Osgouei, and Choi[17], users negatively notice even a 40ms delay in a touchscreen event-based application with haptic feedback. By introducing a network and considering sensory effects metadata processing time in mulsemedia systems, a set of other hurdles emerge, such as network delay, jitter, packet loss, sensory effects metadata transmuting, timetable building, among others, which can have an impact on the quality of experience of users. Therefore, timing issues within the integration of different applications to deliver sensory effects must be considered so as they do not affect performance.

The integration of a gestural interactive and a distributed mulsemedia systems proposed by Santos, Neto, and Saleme[6] sparked the question of how long it took since the gesture recognition until the delivery of the sensory effects to the users. Saleme and Santos[5] showed evidence that the decoupled approach did not cause relevant delays and it has nearly no synchronization loss with timeline applications such as movies. However, this work did not show how long the computational time takes while integrating an interactive application based on events such as those generated when gestures are recognized with sensory effects. Saleme, Celestrini, and Santos[7] timely argued that it was important to know the integration time because future integrations might take advantage of that to improve their architectures. Not only did they show the time for the integration, they also highlighted four major aspects needed to improve response time interval in networked event-based mulsemedia systems: (i) repetitive actions of sending metadata scripts in real time; (ii) high number of exchanged messages; (iii) use of wireless network; and (iv) use of suitable programming languages for processing sensory effect metadata described in MPEG-V. Aspects (i) and (ii) were subsequently tackled by Saleme, Santos, and Ghinea[8].

In their study, Saleme, Santos, and Ghinea[8] reshaped the communication strategy of the PlaySEM SER to work more efficiently with other event-based applications, such as games, VR/AR software, and interactive applications. In addition, they added lightweight communication protocols in its interface to analyze whether they reduce network overhead. To carry out the experiment, they developed mock applications for different protocols to simulate an interactive application working with the PlaySEM SER, measuring the delay between them. The results showed that by preprocessing sensory effects metadata before real-time communication and selecting the appropriate protocol, response time interval in networked event-based mulsemedia systems can decrease remarkably. Communication protocols such as CoAP and MQTT tend to deliver a response in a quite short time with the former being twice as fast as the latter. Compared to CoAP, UPnP was 4 times slower, but it still remains as an option working around 8ms. WebSocket hit 51ms on average. If on the one hand, it makes the process of integration easier for web applications, on the other hand, it could be infeasible for them to work efficiently.

Quality of Experience Assessment Using the PlaySEM Platform

Jalal et al.[18] assessed the quality of experience of users using the PlaySEM SER as a tool to compose their IoT-based architecture for mulsemedia delivery to TV users in a home entertainment scenario. The authors champion that one of the targets of the next generation of TV broadcast services is to provide realistic media content to the users. Not only did they use the PlaySEM SER, they also ran the PlaySEM SE Video Player created by Saleme and Santos[5]. Both solutions were placed respectively in the *Aggregation Layer* and *Application Layer* of their IoT-based architecture, which is composed of two more layers: *Virtualization* and *Physical*.

The environment assembled by Jalal et al.[18] included three RGB Smart Philips LED light devices for lighting effect, an air conditioner for air flow, and a smartphone for vibration effects. The assessment involved 40 participants, 2 per session, to give feedback about a video sequence enriched with mulsemedia. The results pointed that 85% of the users agreed that sensory effects

increase the sense of reality in the experiment, 67.5% were not distracted by the sensory effects, 80% enjoyed the experience with mulsemedia, and 70% judged appropriated the timing of the sensory effects. This last case study showed the flexibility of the PlaySEM SER to work with other devices in a different setup from the previous case studies. Furthermore, it exposed the tool to user evaluations having a commendatory feedback.

CONCLUSION

To shed light on how to cope with the current challenges of delivering mulsemedia described here we summoned up the PlaySEM SER and its case studies. It deals with heterogeneity at several levels such as the diversity of communication protocols, sensory effects metadata standardization, connectivity protocols, and devices. Nonetheless, the plethora of ways to interact with multimedia systems such as multi-haptic, brain, and voice interfaces, added to wearables sensors and actuators spread through the body might pose forthcoming issues in this area. For instance, suppose that the haptic vest worn by the kids in the chemistry experience had at least 100 different points of vibration with different levels of intensity throughout the kid's limbs and trunk. In this case, many concurrent actuators with fine-grained sensory effects might create a delivery bottleneck. Moreover, the perception of heat envisaged in the same experience gives rise to a hitherto not dealt problem of context-awareness in mulsemedia applications. Would the kids feel the same experience in environments with distinct temperatures? A solution to calibrate the environment before delivering the sensory effects by capturing its state through sensors in execution time might be conceived.

The upsurge of commercial, off-the-shelf devices to produce such novel user sensations, erstwhile unimaginable, combined with immersive first-person devices might lead to the development of truly immersive mulsemedia experiences. The flourishing of this new kind of experience will ultimately rely on the user acceptance, which in turn, permeates not only the levels of enjoyment that they feel, but also the way that technology delivers mulsemedia.

ACKNOWLEDGMENTS

REFERENCES

1. H. Purchase, "Defining multimedia," *IEEE MultiMedia*, vol. 5, no. 1, 1998, 8 pages.
2. Y. Arens, E. Hovy, and M. Vossers, "On the knowledge underlying multimedia presentations," *In Intelligent multimedia interfaces*, Mark T. Maybury (Ed.), American Association for Artificial Intelligence, Menlo Park, CA, USA, 1993, 280-306.
3. T. Nakamoto, S. Ito, S. Kato, and G. P. Qi, "Multicomponent Olfactory Display Using Solenoid Valves and SAW Atomizer and its Blending-Capability Evaluation," *IEEE Sensors Journal*, vol. 18, no. 13, 2018, 6 pages.
4. J. A. Covington, S. O. Agbroko, and A. Tiele, "Development of a Portable, Multichannel Olfactory Display Transducer," *IEEE Sensors Journal*, vol. 18, no. 12, 2018, 6 pages.
5. E. B. Saleme and C. A. S. Santos, "PlaySEM: a Platform for Rendering MulSeMedia Compatible with MPEG-V," *In Proceedings of the 21st Brazilian Symposium on Multimedia and the Web*, ACM Press, New York, New York, USA, 2015, 145–148.

6.   C. A. S. Santos, A. N. R. Neto, and E. B. Saleme, "An Event Driven Approach for Integrating Multi-sensory effects to Interactive Environments," *In 2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015, 981–986.

7.   E. B. Saleme, J. R. Celestrini, and C. A. S. Santos, "Time Evaluation for the Integration of a Gestural Interactive Application with a Distributed Mulsemedia Platform," *In Proceedings of the 8th ACM on Multimedia Systems Conference*, ACM Press, New York, NY, USA, 2017, 308-314.

8.   E. B. Saleme, C. A. S. Santos, and G. Ghinea, "Improving response time interval in networked event-based mulsemedia systems," *In Proceedings of the 9th ACM Multimedia Systems Conference*, Amsterdam, Netherlands. ACM, New York, NY, USA, 2018, 216-224.

9.   G. Ghinea, C. Timmerer, W. Lin, and S. R. Gulliver, "Mulsemedia: State of the Art, Perspectives, and Challenges," *ACM Trans. Multimedia Comput. Commun. Appl.* 11, 1s, Article 17, 2014, 23 pages.

10. N. Murray, O. A. Ademoye, G. Ghinea, and G-M. Muntean, "A Tutorial for Olfaction-Based Multisensorial Media Application Design and Evaluation," *ACM Comp. Surveys* 50, 5, Article 67 (2017), 2017, 30 pages.

11. M. Montagud, P. Cesar, F. Boronat, and J. Jansen, "MediaSync: Handbook on Multimedia Synchronization," Springer, 2018.

12. R. T. Fielding, "Architectural styles and the design of network-based software architectures," Doctoral dissertation, University of California, Irvine, 2000.

13. H. Nakamura and H. Miyashita, "Development and Evaluation of Interactive System for Synchronizing Electric Taste and Visual Content," *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, 2012, 517–520.

14. N. Murray, B. Lee, Y. Qiao, and G. Miro-Muntean, "The Impact of Scent Type on Olfaction-Enhanced Multimedia Quality of Experience," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 9, 2017, 13 pages.

15. Z. Yuan, T. Bi, G-M. Muntean, and G. Ghinea, "Perceived Synchronization of Mulsemedia Services," *IEEE Transactions on Multimedia,* vol 17, no. 7, 2015, 10 pages.

16. B. D. Adelstein, T. G. Lee, S. R. Ellis, "Head tracking latency in virtual environments: psychophysics and a model", *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* SAGE Publications, Los Angeles, CA, 2003, pp. 2083-2087.

17. J. R. Kim, R. H. Osgouei, and S. Choi, "Effects of visual and haptic latency on touchscreen interaction: A case study using painting task," *In IEEE World Haptics Conference*, Munich, 2017, pp. 159-164.

18. L. Jalal, M. Anedda, V. Popescu, and M. Murroni, "QoE Assessment for IoT-Based Multi-Sensorial Media Broadcasting," *IEEE Transactions on Broadcasting*, 2018, 9 pages.

ABOUT THE AUTHORS

Estêvão Bissoli Saleme is a doctorate candidate in the Computer Science Department at Federal University of Espírito Santo, Brazil, and an Academic Visitor at Brunel University, United Kingdom. He received the graduation degree in Information Systems from FAESA, Brazil, in 2008, the postgraduate degree in Software Engineering from the Federal University of Lavras, Brazil, in 2010, and the M.Sc. degree in Computer Science from the Federal University of Espírito Santo, Brazil, in 2015. His research interests include multimedia/mulsemedia systems, middlewares and frameworks, interactive multimedia, media transport and delivery. Contact him at estevaobissoli@gmail.com.

Dr. Celso A. S. Santos is a Professor in the Department of Informatics at Federal University of Espírito Santo, Brazil. He received the B.S. degree in Electrical Engineering from the Federal University of Espírito Santo in 1991, and the M.S. degree in Electrical Engineering (Electronic Systems) from the University of São Paulo, São Paulo-SP, Brazil, in 1994. In 1999, he received his Dr. degree at Informatique Fondamentalle et Parallelisme from Université Paul Sabatier de Toulouse III, Toulouse, France. His recent research interests focus on multimedia

systems and applications, synchronization, digital video, and crowdsourcing systems. Contact him at saibel@inf.ufes.br.

Dr. Gheorghita Ghinea is a Professor in the Computer Science Department at Brunel University, United Kingdom. He received the B.Sc. and B.Sc. (Hons) degrees in Computer Science and Mathematics, in 1993 and 1994, respectively, and the M.Sc. degree in Computer Science, in 1996, from the University of the Witwatersrand, Johannesburg, South Africa; he then received the Ph.D. degree in Computer Science from the University of Reading, United Kingdom, in 2000. His work focuses on building adaptable cross-layer end-to-end communication systems incorporating user multisensorial and perceptual requirements. He is a member of the IEEE and the British Computer Society. Contact him at george.ghinea@brunel.ac.uk.