Desafios da modelagem de aplicações multimídia com múltiplos efeitos sensoriais

Douglas P. Mattos[§] Fábio Barreto[§†] Glauco F. Amorim[*§]
Joel A. F. dos Santos[*§] Débora C. Muchaluat-Saade[§]
[*] EIC - CEFET/RJ
[§] Laboratório MídiaCom - UFF
[†] UNILASALLE/RJ
(douglas, fabio, gamorim, joel, debora)@midiacom.uff.br

ABSTRACT

Multimedia applications are usually composed by audiovisual content. Multiple sensorial media, or mulsemedia, applications consider the use of sensorial effects that can stimulate touch, smell and taste, in addition to hearing and sight. Traditional multimedia conceptual models, and consequently multimedia authoring declarative languages, which are used for representing multimedia applications, do not support the definition of multiple sensorial effects. This paper discusses new issues and requirements brought by mulsemedia applications that should be considered in the definition of a mulsemedia conceptual model. A new model must be defined in order to extend multimedia authoring languages to be capable of representing mulsemedia content.

Keywords

MulSemedia, MEPG-V, Sincronização Espaço-Temporal, Modelo Baseado em Eventos

1. INTRODUÇÃO

Considerando o tipo de conteúdo apresentado ao usuário, as aplicações multimídia tradicionais envolvem apenas dois dos sentidos humanos: visão e audição. Entretanto, de acordo com [4], 60% da comunicação humana não é verbal, sendo que a maioria de nós percebe o mundo utilizando cinco sentidos (visão, audição, tato, paladar e olfato). Assim, diferentes trabalhos publicados na literatura propõem o uso de efeitos sensoriais em aplicações multimídia com o intuito de proporcionar novas sensações aos usuários durante a apresentação de uma aplicação multimídia.

As chamadas aplicações mulsemídia [4] (*MulSemedia - multiple sensorial media*), envolvem a apresentação de itens de informação que compreendem não só os sentidos de visão e audição, mas também o tato, o paladar e o olfato. Desta forma, é possível intensificar a sensação de imersão do usuário na aplicação, melhorando assim sua qualidade de experiência (*Quality of Experience* - QoE) [12, 10, 16].

Efeitos sensoriais são percebidos através da modificação das características de um ambiente. Um exemplo de aplicação mulsemídia são os chamados "cinemas 4D", onde cadeiras de movimento são sincronizadas com o conteúdo audiovisual. Um outro exemplo ocorre em jogos digitais, onde a adição de múltiplos efeitos sensoriais podem melhorar a imersão dos usuários no ambiente simulado amplificando a sensação de realidade dos jogos. Em [4], é discutida a aplicação dos efeitos sensoriais na área de saúde, especificamente em seu uso terapêutico, para pessoas com necessidades especiais. Todas estas aplicações visam inserir o usuário em ambientes imersivos para aumentar a qualidade de experiência do usuário em relação à visualização e interação com o conteúdo multimídia apresentado.

O grupo de padronização internacional MPEG, através da especificação MPEG-V [5], especifica um padrão para troca de dados entre o *mundo real*, ambiente onde a aplicação é executada e consumida pelo usuário, e o *mundo virtual*, isto é, a aplicação em si. Tal troca de informações é baseada em esquemas XML, que especificam efeitos sensoriais a serem *executados* no ambiente, bem como informações obtidas de sensores, as descrições dos dispositivos presentes no ambiente, suas capacidades e preferências do usuário.

Aplicações multimídia são comumente especificadas de maneira textual, utilizando algum tipo de linguagem de autoria. Esta especificação pode seguir um paradigma procedural ou declarativo. No paradigma procedural, um programa ou *script* define todos os passos necessários para a apresentação das mídias da aplicação, inclusive sua sincronização. No paradigma declarativo, linguagens de autoria provêem construções em um alto nível de abstração para a especificação das mídias presentes em uma aplicação e a sincronização entre elas.

A ideia principal de uma linguagem de autoria declarativa é separar a descrição de uma aplicação das especificidades de sua execução [6]. Dessa forma, futuras evoluções na forma de apresentação de uma aplicação não impactam na redefinição, ou atualização, de toda uma base de aplicações previamente especificadas. Exemplos de linguagens de autoria multimídia são HTML5 [15], NCL (*Nested Context Language*) [7] e SMIL (*Synchronized Multimedia Integration Language*) [14].

Além dos benefícios citados anteriormente, um outro benefício das linguagens de autoria declarativas, no ambiente multimídia, é a facilidade de autoria de aplicações multimídia. Tal facilidade é importante visto que aplicações multimídia podem ser utilizadas em diferentes áreas, como web, TV digital e IPTV; e por diferentes perfis de autores, como desenvolvedores e produtores de conteúdo.

In: Workshop Internacional de Sincronismo das Coisas (WSoT), 1., 2016, Teresina. Anais do XXII Simpósio Brasileiro de Sistemas Multimídia e Web. Porto Alegre: Sociedade Brasileira de Computação, 2016. v. 2.

ISBN: 978-85-7669-332-1

WebMedia'16, November, 2016, Teresina, Piauí, Brazil

Apesar do MPEG-V especificar um padrão de comunicação entre a aplicação e o ambiente no qual ela é executada, bem como uma gama de efeitos sensoriais, a definição da aplicação é deixada a cargo do autor. Assim, é necessário especificar os tipos de mídias e efeitos a serem apresentados, sua sincronização ao longo do tempo e como o usuário poderá interagir com a aplicação. Visando facilitar a autoria de tais aplicações, trabalhos publicados na literatura propõem ferramentas de autoria mulsemídia [2, 8, 13], abstraindo parte do código da aplicação do autor. Apesar de facilitar a autoria de tais aplicações, estas ferramentas não provêem uma clara separação entre a aplicação e as especificidades de sua apresentação, como pode ser observado no paradigma de autoria declarativo.

Este artigo discute desafios sobre a modelagem de aplicações mulsemídia. A ideia principal é permitir a especificação de uma aplicação contendo múltiplos efeitos sensoriais de maneira similar à especificação de uma aplicação multimídia tradicional. Tal proposta é alcançada através da extensão da modelagem conceitual usada em aplicações multimídia, focando não só no conteúdo audiovisual apresentado ao usuário, mas em efeitos sensoriais apresentados em um ambiente imersivo. Conforme, será discutido, é necessário prover um novo modelo conceitual, estendendo modelos multimídia existentes, para a descrição completa de uma aplicação mulsemídia.

O restante desse artigo está estruturado como segue. A Seção 2 apresenta brevemente o padrão MPEG-V, sua arquitetura e efeitos sensoriais previstos, comentando suas limitações para representação da sincronização espaço-temporal. A Seção 3 discute desafios da modelagem de aplicações mulsemídia. A Seção 4 conclui esse artigo, apresentando alguns trabalhos futuros.

2. MPEG-V

O padrão MPEG-V define elementos baseados em XML para especificar objetos do mundo real (sensores e atuadores) e virtual (objetos virtuais) permitindo a troca padronizada de dados entre esses mundos. Este padrão é dividido em 7 partes e cada uma delas será discutida brevemente a seguir.

A Parte 1 [5] descreve a arquitetura do padrão e a integração das suas partes, visando garantir a interoperabilidade entre o mundo real e virtual. A arquitetura do MPEG-V pode ser visualizada na Figura 1.

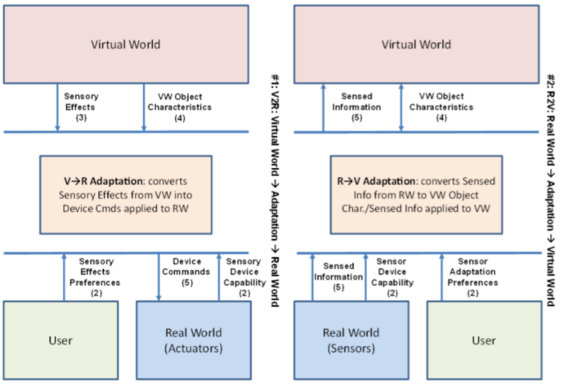

Figure 1: Arquitetura do padrão MPEG-V. [5]

A parte esquerda da arquitetura corresponde ao mapeamento do mundo virtual para o real (V2R), onde efeitos sensoriais (Parte 3) e características de objetos virtuais (Parte 4) são reproduzidos no mundo real. Para isso, o adaptador (que não faz parte do padrão) deve transformar os efeitos sensoriais em comandos para os dispositivos (Parte 5) e executá-los conforme preferências dos usuários (Parte 2). Além disso, ele deve ser capaz de coletar informações sobre a capacidade dos atuadores (Parte 2) presentes no mundo real. Note que atuadores podem ser compatíveis diretamente com o padrão MPEG-V não necessitando de realizar a tradução dos comandos enviados.

Já a parte da direita da arquitetura apresentada na Figura 1 realiza o mapeamento do mundo real para o virtual (R2V), obtendo as informações dos sensores (Parte 5), da capacidade dos sensores (Parte 2) e preferências do usuário em relação aos sensores (Parte 2). Com essas informações, o adaptador deve reproduzi-las no ambiente virtual. Através das características dos objetos virtuais (Parte 4), é possível trocar informações entre diferentes mundos virtuais. Regras de conformidade e software de referência para validar e esclarecer as especificações descritas com as outras partes do padrão são definidas na Parte 7.

Os parágrafos a seguir discutem brevemente a especificação de efeitos sensoriais, objetos virtuais e informações obtidas de sendores contida nas partes 3, 4 e 5 do padrão.

A Parte 3 especifica uma linguagem denominada SEDL (*Sensory Effect Description Language*) para descrever os efeitos sensoriais que serão transformados pelo adaptador VR para comandos nos atuadores presentes no mundo real. Ela possui um vocabulário denominado SEV (*Sensory Effect Vocabulary*) que define tipos de efeitos sensoriais e também permite que o autor especifique outros. Um arquivo de efeitos sensoriais especificados com a linguagem SEDL, denominado *Sensory Effect Metadata* (SEM), associa o efeito a algum conteúdo multimídia. O padrão ainda define um modelo espacial para descrever a localização a partir da qual os efeitos sensoriais são recebidos a partir da perspectiva do usuário, permitindo especificá-los para uma determinada região do ambiente (*back*, *midway*, *front* etc).

Além disso, a Parte 3 do MPEG-V define um modelo temporal, em que cada efeito sensorial pode ser ativado e desativado em um determinado instante de tempo através de atributos da linguagem SEDL. Nesse modelo, ainda podem ser especificados intervalos de tempo denominados *fade-in* e *fade-out*, nos quais a intensidade definida para o efeito sensorial deve ser alcançada ou deve atingir o valor nulo respectivamente.

A Parte 4 permite descrever os objetos virtuais, além disso ela define a sintaxe e semântica da descrição destes objetos permitindo que eles sejam controlados por entradas originadas do mundo real e sejam reutilizados em outros mundos virtuais. Os tipos básicos de atributos definidos são representados por: identidade (contém um descritor de identificação), som (recursos de som), aroma (recursos de aroma), controle (descritores para controlar os movimentos do objeto (translação, orientação e redimensionamento)), eventos (eventos de entrada), modelo de comportamento e um ID para identificar cada objeto virtual.

Vale ressaltar que esta parte do MPEG-V define características dos objetos (metadados) e não especifica formas geométricas, animação ou textura. Para suportar estas especificações e oferecer uma solução de interoperabilidade completa, as Partes 16 e 11 do padrão MPEG-4 [17] podem ser

utilizadas com o MPEG-V. A Parte 16 do MPEG-4 inclui um framework para definir e animar avatares e a Parte 11 auxilia na definição de elementos gráficos.

A Parte 5 define a linguagem IIDL (*Interaction Interface Description Language*) para permitir a descrição dos comandos dos atuadores (V2R) e de informações obtidas dos sensores (R2V). Esses comandos definidos com a linguagem IIDL são mais simples do que os especificados com a Parte 3 do padrão, assim a Parte 5 busca facilitar a implementação da interpretação dos comandos nos dispositivos.

No padrão MPEG-V, a sincronização de um evento sensorial com um conteúdo audiovisual é feita através de arquivos SEM enviados ao adaptador. A sincronização definida nesses arquivos é feita através da especificação de instantes de tempo, quando um efeito deve ser realizado. Assim, a sincronização prevista pelo MPEG-V segue um modelo baseado em eixo temporal.

Como discutido em [1], tal abordagem não é muito expressiva. Portanto, é importante definir um modelo conceitual para a descrição de aplicações mulsemídia. A Seção a seguir discute os requisitos e desafios por trás de especificação de tal modelo.

3. MODELANDO APLICAÇÕES MULSEMEDIA

Em uma aplicação multimídia tradicional, toda informação que é efetivamente consumida pelo usuário, como um vídeo, um áudio, uma imagem e etc pode ser representada como um nó ou objeto de mídia. Nós são entidades cuja representação do conteúdo é abstraída na descrição da aplicação. Um nó pode representar uma mídia de texto, áudio, vídeo, imagem, um script ou um programa. O conteúdo de um nó é composto de um conjunto de unidades de informação dependentes do tipo de mídia (quadros de um vídeo, amostras de um áudio, pixels em uma imagem, caracteres em um texto, etc). Uma âncora de um nó representa uma parte, um subconjunto de seu conteúdo.

Já as relações entre tais mídias podem levar em consideração a ocorrência de eventos, como por exemplo: interação do usuário ou o resultado de uma computação (um script auxiliar). A maneira como as relações são definidas depende do modelo de sincronização temporal no qual a linguagem de autoria é baseada. As entidades dos modelos conceituais que representam relações são os elos e as composições [22]. Elos são comumente usados para representar relações de referência, interatividade ou derivação de versões entre nós. Por sua vez, as composições podem representar diferentes tipos de relação, como por exemplo, estruturação lógica ou sincronização temporal entre seus componentes.

A exibição de um nó, em uma aplicação multimídia, pode induzir um intervalo temporal representando o período ao longo da apresentação da aplicação em que é o nó é exibido. É possível que tal intervalo se repita um número indefinido de vezes durante a aplicação. A definição do tamanho, posição e repetição (no tempo) deste intervalo é comumente definido por relações especificadas na aplicação. Ainda, um nó durante sua exibição ocupa um determinado canal, por exemplo, um canal de áudio ou um espaço na tela do dispositivo de exibição. O local onde um nó é apresentado é comumente estabelecido através de seu posicionamento em relação a tela do dispositivo de exibição, de maneira absoluta ou relativa. Em alguns casos, é possível ainda que o posicionamento de um nó seja definido através de relações espaciais (geralmente através de restrições) [20].

Visando representar uma aplicação mulsemídia, é necessário definir extensões dos modelos conceituais multimídia tradicionais, que possam reproduzir estímulos sensoriais.

Uma discussão inicial importante envolve a definição de requisitos que este novo tipo de modelo deve satisfazer. Pode-se destacar os seguintes requisitos: i) suporte para representar o conteúdo de diferentes tipos de mídias e estímulos sensoriais; ii) suporte a multiusuário e interação multimodal, que considera por exemplo, interface via voz, gesto, toque, etc.; iii) suporte a entrada de dados através de dispositivos de entrada e/ou sensores; iv) suporte a exibição de dados através de dispositivos de saída e/ou atuadores; v) sincronização temporal e espacial entre mídias e diferentes atuadores e sensores; vi) suporte à definição de relações de restrição; vii) suporte para representar informações de estado; viii) adaptação de conteúdo conforme o perfil do usuário e ix) adaptação do conteúdo conforme características do ambiente e dos dispositivos de exibição.

Levando-se em conta os requisitos descritos, um modelo mulsemídia deve contemplar a definição de relações de sincronização entre os componentes de uma aplicação, permitindo integrar exibição de conteúdo, interatividade dos usuários e mudanças de informações de estado da aplicação. Um modelo conceitual baseado em eventos, tal como o NCM – *Nested Context Model* [21], tem potencial para oferecer tais facilidades. Entretanto, a definição desse novo modelo mulsemídia não é trivial, devendo ser consideradas as seguintes questões.

Um dado estímulo sensorial possui uma representação no tempo? Ele pode ser representado simplesmente como um intervalo, representando os instantes de tempo onde aquele estímulo está presente no ambiente? Dependendo do estímulo, devemos considerar o tempo até que o estímulo possa ser percebido na intensidade desejada? Por exemplo, supondo que um determinado estímulo altera a temperatura ambiente, essa alteração será considerada somente quando a temperatura atingir a desejada?

Um dado estímulo sensorial ocupa algum tipo de canal? Um áudio ocupa um canal de áudio, enquanto um vídeo ocupa uma parte da tela do dispositivo de exibição. Os demais estímulos sensoriais ocupam, de maneira similar, algum canal? Podemos considerar o próprio ambiente real como o canal a ser ocupado?

Pode haver sobreposição de estímulos do mesmo tipo? É possível que dois estímulos do mesmo sentido se sobreponham? Por exemplo dois ou mais cheiros podem ser "exibidos" simultaneamente?

Uma outra questão importante na definição de tal modelo é se a representação de um dado estímulo no tempo e no espaço (como feito em objetos multimídia tradicionais) é suficiente para a sua especificação? Seria necessário algum outro tipo de eixo para a representação de um estímulo sensorial. Modelos de sincronização multimídia são fortemente associados ao tempo, no sentido que toda a sincronização é feita de modo que uma dada mídia seja apresentada em um determinado momento. Na descrição de uma aplicação mulsemídia, o tempo pode não ser o fator mais importante. Por exemplo, suponha que deseja-se modificar a configuração de um ambiente de acordo com a posição do espectador dentro dele. Nesse caso, o tempo não é o fator mais importante e sim a posição espacial. Seria a posição do usuário

um terceiro eixo a ser considerado na especificação de uma aplicação? Repensar nos modelos de sincronização multimídia existentes a luz dessas características é um dos grandes desafios atuais.

Em documentos multimídia, entrada de dados pelo usuário são modeladas através da interação, como seleção de mídias, ou mudança de valores de atributos das mídias. Dada a possibilidade do uso de sensores em aplicações mulsemídia, outras formas de interação com uma aplicação tornam-se possíveis. Exemplos são a interação por gestos, comandos verbais, expressões faciais (ou corporais), diferentes tipos de toque (*touch, force touch*) etc. Tais formas de interação devem ser consideradas na modelagem de tais aplicações, inclusive seu uso para sincronização entre estímulos sensoriais.

4. CONCLUSÃO

Este artigo discutiu conceitos que envolvem as aplicações com múltiplos efeitos sensoriais e apresentou o padrão MPEG-V, que auxilia na troca de informações entre o mundo real e virtual. Este padrão fornece diversas linguagens baseadas em XML para especificar os dispositivos presentes no mundo real (ex.: atuadores e sensores) descrevendo suas capacidades e comandos a serem executados por eles, além de permitir definir de forma padronizada efeitos sensoriais e os objetos virtuais, que podem ser renderizados para o mundo real.

A área pesquisa em mulsemídia possui ainda diversos desafios para alavancar novas aplicações que os efeitos sensoriais podem proporcionar. Um destes desafios é a integração efetiva dos efeitos sensoriais ao conteúdo audiovisual das aplicações multimídia tradicionais, implicando a definição de um novo modelo mulsemídia que dê suporte aos requisitos descritos na Seção 3. Como trabalhos futuros, pretende-se estender o modelo NCM para transformá-lo em um modelo mulsemídia. Além disso, novas ferramentas de autoria mulsemídia podem auxiliar na definição dessas novas facilidades e precisam ser desenvolvidas. Novos estudos em QoE para aplicações com múltiplos efeitos sensoriais também devem ser realizados.

5. REFERENCES

[1] G. Blakowski and R. Steinmetz, "A media synchronization survey: Reference model, specification and case studies," *Journal on Selected Areas in Communications*, vol. 14, no. 1, pp. 5–35, janeiro 1996.

[2] Choi, B., Lee, E. S., and Yoon, K. (2011). "Streaming media with sensory effect". In 2011 International Conference on Information Science and Applications.

[3] Choi, B. S.; Kim, S. K. (2012). "Text of ISO/IEC FDIS 23005-3 2nd edition sensory information".

[4] Ghinea, G., Timmerer, C., Lin, W., and Gulliver, S. R. (2014). "Mulsemedia: State of the art, perspectives, and challenges". *ACM Trans. Multimedia Comput. Commun. Appl.*

[5] Han, J. J. and Kim, S. (2014). "Text of white paper on MPEG-V". In MPEG Communication Group, 107.

[6] Hardman, H. L. "Modeling and Authoring Hypermedia Documents". (1998). Ph.D. Thesis — Universität Amsterdam.

[7] "Nested Context Language (NCL) and Ginga-NCL for IPTV services". (2009). http://www.itu.int/rec/T-REC-H.761-200904-S. ITU-T Recommendation H.761.

[8] Kim, S. K. (2013). "Authoring multisensorial content". Signal Processing: Image Communication.

[9] Kim, S. K. and Preda, M. (2014). "MPEG-V international standard and the internet of things". In International Conference on Software Intelligence Technologies and Applications International Conference on Frontiers of Internet of Things 2014.

[10] Rainer, B., Waltl, M., Cheng, E., Shujau, M., Timmerer, C., Davis, S., Burnett, I., Ritz, C., and Hellwagner, H. (2012). "Investigating the impact of sensory effects on the quality of experience and emotional response in web videos". In International Workshop on Quality of Multimedia Experience (QoMEX)

[11] L. F. G. Soares and R. F. Rodrigues, "Nested context model 3.0 part 1 - ncm core," Tech. Rep., Informatics Department, PUC-Rio, Rio de Janeiro, maio 2005.

[12] Waltl, M., Timmerer, C., and Hellwagner, H. (2010). "Improving the quality of multimedia experience through sensory effects". In International Workshop on Quality of Multimedia Experience (QoMEX)

[13] Waltl, M., Rainer, B., Timmerer, C., and Hellwagner, H. (2013). "An end-to-end tool chain for sensory experience based on MPEG-V". Signal Processing: Image Communication.

[14] "Synchronized Multimedia Integration Language - SMIL 3.0 Specification". (2008). http://www.w3c.org/TR/SMIL3. World-Wide Web Consortium Recommendation.

[15] "HTML5: A vocabulary and associated APIs for HTML and XHTML". (2014). http://www.w3.org/TR/html5. World-Wide Web Consortium Recommendation.

[16] Yuan, Z., Ghinea, G., and Muntean, G. M. (2015). "Beyond multimedia adaptation: Quality of experience-aware multi-sensorial media delivery". IEEE Transactions on Multimedia.

[17] Richardson, Iain E. H. "264 and MPEG-4 video compression: video coding for next-generation multimedia". John Wiley & Sons, 2004.

[18] Concolato, C., Cordara, G., Park, K. (2013). "Text of ISO/IEC 23007: MPEG Rich Media User Interface (MPEG-U)".

[19] Martínez, J. M. (2003). "Text of ISO/IEC 15938-5:2003, Information Technology – Multimedia Content Description Interface – Part 5: Multimedia Description Schemes".

[20] Amorim, G. F., dos Santos, J. A. F., and Muchaluat-Saade, D. C. (2016). "STyLe: Extending NCL for providing Dynamic Layouts". will appear in WebMedia 2016.

[21] Soares, L. F. G. and Rodrigues, R. F. (2005). "Nested Context Model 3.0: Part 1 – NCM Core". ISSN: 0103-9741.

[22] Soares, L. F. G., Rodrigues, R. and Muchaluat-Saade, D. C. (2000). "Modeling, authoring and formatting hypermedia documents in the HyperProp system". Multimedia Systems, Volume 8, Number 2, Page 118.