

Preparação de Objetos de Mídia e Efeitos Sensoriais para Formatação de Documentos Mulsemídia

Marina Josué
marinaivanov@midia.com.uff.br
Universidade Federal Fluminense
Niterói, Rio de Janeiro, Brasil

Débora Muchaluat Saade
debora@midia.com.uff.br
Universidade Federal Fluminense
Niterói, Rio de Janeiro, Brasil

Marcelo Moreno
moreno@ice.ufjf.br
Universidade Federal de Juiz de Fora
Juiz de Fora, Minas Gerais, Brasil

Resumo

Aplicações mulsemídia permitem a incorporação de efeitos sensoriais à multimídia tradicional visando aumentar a QoE dos usuários. Na fase de execução, essas aplicações podem sofrer com atrasos na entrega do conteúdo ou durante a renderização de um efeito sensorial. Esses atrasos podem levar a falhas de sincronização durante a apresentação do conteúdo. Neste contexto, este artigo apresenta a proposta da tese de doutorado que visa evitar tais falhas por meio de uma nova operação para preparação de objetos de mídia e efeitos sensoriais.

Keywords: Aplicação Mulsemídia, Preparação de Objetos de Mídia, Preparação de Efeitos Sensoriais, Formatação Mulsemídia, Aplicações NCL, Ginga-NCL.

1 Introdução

Aplicações multimídia tradicionais podem ser enriquecidas com efeitos sensoriais para fornecer experiências imersivas com maior qualidade aos usuários. Essas aplicações são denominadas aplicações mulsemídia [2], e permitem estabelecer relações de sincronização temporal entre o conteúdo audiovisual e efeitos sensoriais, que estimulam outros sentidos como tato, olfato e paladar. Durante a fase de transmissão, o conteúdo das aplicações pode ser transmitido utilizando uma rede de comunicação, e por isso, a entrega dos dados pode sofrer com atrasos devido a problemas de congestionamento da rede. Além dos objetos de mídia tradicionais, a apresentação de efeitos sensoriais pode também sofrer com atrasos, devido à limitação de capacidade dos dispositivos responsáveis pela renderização dos efeitos, e do tempo de comunicação entre o formatador e o atuador.

Após a entrega da aplicação e de seu conteúdo, o documento mulsemídia é interpretado e executado em um ambiente de apresentação que fornece suporte às funcionalidades fornecidas pela linguagem de autoria. Esta fase também é conhecida como fase de formatação da aplicação, e é desempenhada por um componente chamado formatador (também

conhecido como máquina de apresentação, ou do inglês, *presentation engine*) [8]. O formatador é responsável por funções como carregamento e *parsing* do documento multimídia, acionamento de *players* de mídia e controle das interações do usuário com a aplicação. Além disso, nos sistemas mulsemídia, o formatador também deve ser capaz de se comunicar com os dispositivos atuadores responsáveis pela renderização dos efeitos, e controlar a sincronização de efeitos sensoriais entre si, e entre efeito sensorial e objetos de mídia.

Um dos objetivos da tese de doutorado [6] foi prover mecanismos para tentar reduzir ou até mesmo evitar falhas de sincronização durante a apresentação de aplicações mulsemídia. Além disso, o trabalho de tese propôs modificações na linguagem NCL (*Nested Context Language*) [1] e no modelo conceitual NCM (*Nested Context Model*) para dar suporte à especificação de aplicações com efeitos sensoriais.

2 Objetivos da Tese de Doutorado

Considerando a autoria de aplicações mulsemídia, um dos objetivos da tese foi oferecer um nível mais alto de abstração na definição de efeitos sensoriais em modelos mulsemídia, e conseqüentemente em linguagens de autoria declarativas. Para isso, a tese de doutorado propôs representação de um efeito sensorial como entidade de primeira classe no modelo conceitual, o que permitiu transformar um modelo multimídia em mulsemídia [5]. Desse modo, um efeito sensorial passa a ser definido como um novo tipo de nó, denominado *sensory effect node*, possibilitando que seja especificado em uma aplicação mulsemídia, independente dos dispositivos físicos utilizados para sua renderização.

Um objetivo específico da tese foi incorporar essa representação de efeito na linguagem (NCL), para auxiliar na autoria aplicações mulsemídia no Sistema Brasileiro de TV Digital Terrestre (SBTVD-T) [1], dando origem à versão 4.0 da linguagem. A proposta de NCL 4.0 já foi levada ao Fórum SBTVD como proposta à futura geração da TV digital no Brasil, atendendo ao requisito de TV imersiva¹ e foi avaliada com sucesso pelo Fórum. Portanto, provavelmente será incorporada às normas ABNT do futuro sistema de TVD.

Outro objetivo da tese foi auxiliar no controle da sincronização de aplicações multimídia e mulsemídia na fase de formatação. Neste contexto, a tese propôs um novo tipo de

In: IV Concurso de Teses e Dissertações (CTD 2022), Curitiba, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2022.

© 2022 SBC – Sociedade Brasileira de Computação.
ISSN 2596-1683

¹https://forumsbtvd.org.br/tv3_0/

operação, denominado preparação, que pode ser aplicado tanto a objetos de mídia quanto a efeitos sensoriais, a fim de reduzir, ou até mesmo evitar falhas de sincronização durante a apresentação do conteúdo.

3 Contribuições da Tese

As contribuições da tese são divididas em duas frentes, que serão discutidas a seguir.

3.1 Autoria de aplicações mulsemídia

Em relação à fase de autoria de aplicações mulsemídia, a tese possui contribuições tanto na modelagem de aplicações, quando na linguagem de autoria.

A NCL é uma linguagem declarativa para especificação de aplicações multimídia baseada no modelo conceitual NCM [7]. Dessa forma, para que o NCM também permita a representação de efeitos sensoriais, a tese propôs uma extensão do modelo.

O modelo NCM [7] distingue duas classes básicas de nós, chamados de nós de conteúdo (*content nodes*) e nós de composição (*composite nodes*). Um nó de conteúdo é definido como um fragmento de informação e pode representar um vídeo, áudio, imagem, texto, um conteúdo de script ou até mesmo uma outra aplicação. Já o nó de composição pode ser composto por um ou mais nós de conteúdo, ou até mesmo outros nós de composição.

Baseado na definição de nó presente no modelo, a tese propôs a criação de um novo tipo de nó para representar um efeito sensorial, denominado *EffectNode*. Um nó de efeito sensorial é portanto, uma entidade de primeira classe que abstrai o tipo do efeito empregado e pode ser especializado em diferentes tipos, como efeito de vento (*WindEffectNode*), efeito de temperatura (*TemperatureEffectNode*), efeito de luz (*LightEffectNode*) e efeito de aroma (*ScentEffectNode*), etc. É importante destacar que, ao estender o conceito de um nó NCM para representar efeitos sensoriais, é possível agrupá-los em composições para criar efeitos compostos. Além disso, como o modelo permite a definição de links para interconectar nós, estes podem ser utilizados para representar relações entre efeitos sensoriais, conteúdo multimídia e também envolvendo interações do usuário.

Para que a linguagem dê suporte a especificação de efeitos sensoriais, a tese propôs a criação de um novo módulo funcional, denominado *Effect*. Dessa forma, os efeitos sensoriais são especificados pelo elemento `<effect>`. Isto possibilita a manipulação de efeitos sensoriais de forma similar à manipulação de outros nós em NCL, como objetos de mídia. A Listagem 1 apresenta um exemplo de especificação de um efeito de aroma utilizando o elemento `<effect>` proposto.

```

1 <effect id="flower" type="ScentType"
  descriptor="desc">
2   <property name="scent" value="urn:mpeg:
  mpeg-v:01-SI-ScentCS-NS:rose"/>

```

3 `</effect>`

Listing 1. Definição de efeito sensorial de aroma em NCL.

A fim de permitir identificar e caracterizar a ocorrência de um efeito sensorial foi criado um conjunto de atributos (*id*, *type* e *descriptor*) para o elemento `<effect>`. O atributo *type* permite especificar o tipo do efeito sensorial, enquanto o atributo *descriptor* permite definir propriedades que definem o comportamento de renderização do efeito (e.g. duração, cor do efeito de luz, etc). Por fim, a tese propôs modificações no elemento `<region>`, que especifica a região espacial para renderização do efeito. Com as modificações propostas, é possível especificar o posicionamento do efeito utilizando coordenadas esféricas, ou então o mecanismo de localização proposto no padrão MPEG-V[4].

Em relação à preparação de conteúdo de mídia e efeitos sensoriais, a tese propôs a criação de um novo evento na linguagem, denominado *preparation*. O evento de preparação é genérico o suficiente para ser aplicado tanto a objetos de mídia tradicionais quanto a efeitos sensoriais. O evento de preparação deve ser disparado quando um objeto de mídia tiver seu *player* instanciado e seu conteúdo preparado antecipadamente ao momento de sua apresentação na aplicação, ou quando os dispositivos responsáveis pela renderização de um efeito sensorial devem ser pré-ativados para uma melhor sincronização da percepção desse efeito pelo usuário.

Quando o evento de preparação é aplicado a um objeto de mídia, sua ocorrência chega ao fim quando (i) o buffer do *player* de mídia estiver completo ou (ii) o conteúdo da mídia a ser preparada for completamente carregado pelo *player*; o que acontecer primeiro. Esse comportamento é diferente do evento de pré-busca (*prefetch*), que só finaliza quando todo o conteúdo do objeto de mídia for recuperado a partir de sua origem e não inclui a instanciação do *player* como parte de sua ocorrência. Assim, o evento de preparação é mais viável em certas aplicações multimídia e independente da capacidade de armazenamento do ambiente de reprodução.

A preparação de um efeito sensorial depende do seu tipo. Por exemplo, para o efeito de calor, seria necessário acionar o dispositivo aquecedor antecipadamente para que a temperatura do ambiente chegue ao valor desejado no instante em que deve ser sincronizado com uma cena em um deserto.

Na fase de autoria, o evento de preparação pode ser empregado pelo autor do conteúdo multimídia ou mulsemídia, a fim de programar o preparo de um conteúdo ou efeito sensorial a ser apresentado. Semelhante aos eventos já definidos na linguagem NCL 3.0 (*presentation*, *attribution*, *selection* e *composition*), o evento *preparation* possui os atributos *occurrences* e *repetition*, que podem ser utilizados pelo autor da aplicação ou pelo formatador NCL.

Além dos atributos mencionados, a tese propôs a definição de um novo atributo para o evento de preparação. Este atributo é denominado *prepared*, e permite sinalizar se uma

preparação foi bem sucedida. Através da verificação deste atributo, a reprodução de um nó de mídia do tipo vídeo, por exemplo, pode ser cancelada, caso seu conteúdo não tenha sido preparado, o que pode evitar atrasos de sincronização na aplicação. A proposta do evento de preparação foi levada ao Fórum SBTVD e já foi incorporada à norma ABNT NBR 15606-2 [1] vigente do *middleware* Ginga Perfil D.

3.2 Formatação de aplicações mulsemídia

O Ginga-NCL [1] é o subsistema do *middleware* Ginga responsável por prover uma infraestrutura de apresentação para aplicações especificadas na linguagem NCL. Para possibilitar a execução de aplicações mulsemídia no Ginga-NCL, a tese estendeu a implementação de referência da máquina de apresentação Ginga-NCL².

Para dar suporte à execução de efeitos sensoriais, foi criado um novo componente no Ginga-NCL, denominado *EffectPlayer*, que realiza a comunicação entre o formador e os dispositivos físicos responsáveis pela renderização do efeito. O *player* de efeito é criado quando um efeito tem sua preparação iniciada, e é destruído assim que o efeito é finalizado - através de uma ação de *stop* sobre o efeito, ou quando o mesmo atinge uma duração, definida de forma explícita pelo autor da aplicação.

Cada *player* de efeito está associado a apenas um efeito sensorial, e vice-versa. Entretanto, pode-se ter várias instâncias de *player* de efeito, que se comunicam com um mesmo dispositivo físico. O *player* de efeito pode ser especializado de acordo com o tipo de efeito sensorial que ele manipula.

A máquina de apresentação também foi modificada para dar suporte à preparação automática de efeitos sensoriais e conteúdo de mídia. Além disso, na nova versão implementada na tese, o *middleware* Ginga é capaz de reconhecer o evento *preparation*, especificado pelo autor de uma aplicação NCL para preparações programadas.

O mecanismo de preparação automática de objetos de mídia e efeitos sensoriais tem como ponto inicial a construção do grafo temporal da aplicação. O grafo temporal é composto por vértices que representam transições de estado de um evento, e arestas dirigidas que representam os relacionamentos entre os eventos. Uma aresta no grafo pode ser rotulada com uma condição, que deve ser satisfeita, para que uma ação especificada no vértice de destino da aresta seja disparada.

Durante a fase de autoria de uma aplicação, o autor pode especificar como os objetos de mídia e efeitos sensoriais devem se comportar ao longo tempo, seu posicionamento no espaço, e as relações de sincronização espaço-temporal entre eles. Durante a fase de execução, para garantir uma apresentação de qualidade, é importante que o formador

mantenha essas relações, e para isso, pode ser utilizado um plano de apresentação.

O plano de apresentação é uma estrutura derivada do grafo temporal hipermídia. Ele contém os instantes de ocorrência previsíveis das ações sobre o evento de apresentação de cada objeto de mídia que compõe a aplicação. Desse modo, ao observar um plano de apresentação, é possível verificar o instante no tempo de apresentação em que um objeto inicia (transição *start* no evento *presentation*) e finaliza sua exibição (transição *stop* no evento *presentation*).

Após a criação do plano de apresentação, a tese propõe a criação automática de um plano de preparação por parte do formador, que irá conter os instantes em que cada componente da aplicação multimídia ou mulsemídia deve ser preparado para apresentação. Através deste plano, o formador é capaz de realizar a preparação automática de conteúdo de mídia e efeitos sensoriais, e evitar que ocorram falhas de sincronização durante a apresentação da aplicação. A fim de validar a proposta de preparação automática, a tese implementou a criação dos planos de apresentação e preparação no *middleware* Ginga-NCL. Além disso, foi implementado também a criação do grafo temporal para representar o comportamento temporal da aplicação.

Para a preparação automática de efeitos sensoriais a tese implementa um mecanismo de calibragem de efeitos, possibilitando que a preparação do efeito seja feita considerando as características intrínsecas a cada tipo de efeito e ao ambiente físico em que o *player* é executado.

4 Discussão sobre Resultados Obtidos

O trabalho da tese avaliou a utilização de preparação automática aplicada a conteúdo de mídia tradicional e a efeitos sensoriais. Para a realização dos testes, foi utilizada a versão da implementação de referência do Ginga-NCL estendida pela tese. Os testes de preparação de conteúdo foram realizados em dois ambientes de execução diferentes, que simulam o dispositivo receptor da aplicação, podendo ser um televisor com menos recursos computacionais ou um computador.

Para a avaliação do evento de preparação aplicado a objetos de mídia utilizou-se uma aplicação multimídia de propaganda direcionada. Nessa aplicação, é fornecido ao usuário um conteúdo considerando seu perfil, ao invés do comercial geral distribuído via sinal *broadcast*. Para que este tipo de aplicação possa ser fornecido ao usuário, é necessário que o receptor seja capaz de receber tanto conteúdo enviado por *broadcast* quanto por *broadband*.

Após a realização dos testes foi possível observar que no receptor com menos recursos computacionais, o uso de preparação automática reduziu em 99,8% o atraso para apresentação do conteúdo *broadband* considerando uma taxa de transmissão igual a 8 Mbps. Já nos cenários onde a taxa de transmissão foi igual a 25 Mbps, o atraso para apresentação do conteúdo direcionado sem a preparação automática foi

²<https://github.com/marinaivanov/ginga-mulsemmedia/tree/mulsemmedia-preparation>

98 vezes maior do que o tempo nos casos com preparação automática, nessa mesma taxa.

A análise do evento de preparação aplicado a efeitos sensoriais foi realizada utilizando três aplicações diferentes, uma com efeito de luz, outra com efeito de aroma, e por fim uma aplicação utilizando apenas o efeito de vento. Os testes mostraram que a preparação automática de conteúdo de mídia e efeitos sensoriais reduzem de forma significativa o atraso na apresentação de aplicações mulsemídia. Além disso, foi possível observar que, para efeitos sensoriais que têm sua percepção instantânea por parte do usuário (efeito de vento e de luz, por exemplo), o que impacta o atraso de apresentação é o protocolo de comunicação utilizado pelo atuador.

Já para o efeito de aroma, observou-se um atraso médio de 11,9s para apresentação do efeito quando a preparação automática não foi aplicada. Já nos testes onde a aplicação mulsemídia foi executada utilizando a preparação automática, este atraso médio foi reduzido para 3,84s. De acordo com a análise realizada por [3], o efeito de aroma pode ser renderizado com uma diferença de 5 segundos do trecho de vídeo ao qual está relacionado, sem que os usuários percebam erros de sincronização. Dessa forma, o valor obtido está dentro do intervalo aceitável.

4.1 Publicações

A tese de doutorado apresentada neste artigo gerou as seguintes publicações:

- Josué, M., Abreu, R., Barreto, F., Mattos, D., Amorim, G., dos Santos, J., and Muchaluat-Saade, D. (2018, June). **Modeling sensory effects as first-class entities in multimedia applications**. In Proceedings of the **9th ACM Multimedia Systems Conference – MMSys 2018** (pp. 225-236). ACM, 2018 - Qualis A2.
- Josué, M. P.; Moreno, M. F.; Muchaluat-Saade, D. C. **Preparation of Media Object Presentation and Sensory Effect Rendering in Mulsemimedia Applications**. In: Proceedings of the **24th Brazilian Symposium on Multimedia and the Web - WebMedia'18**, 2018 - Qualis A4.
- Josué, M. P.; Moreno, M. F.; Muchaluat-Saade, D. C. **Mulsemimedia preparation: a new event type for preparing media object presentation and sensory effect rendering**. In: Proceedings of the **10th ACM Multimedia Systems Conference – MMSys 2019**. ACM, 2019. p. 110-120 - Qualis A2.
- Josué, M. I. P.; Moreno, M. F.; Muchaluat-Saade, D. C. **Automatic Preparation of Media Objects in Multimedia Applications**. Proceedings of the **25th Brazilian Symposium on Multimedia and the Web - WebMedia'19**, 2019 - Qualis A4.
- Josué, M.; Montevecchi, E. ; Abreu, R. ; Barreto, F. ; Santos, J. ; Muchaluat-Saade, D. C. **Ambientes multissensoriais aplicados à saúde: desenvolvimento**

de aplicações e tendências futuras. Capítulo do Livro de Minicursos SBCAS 2020. SBC, 2020, p. 48-89.

Além das publicações listadas acima, foram publicados mais 3 artigos em conferências e um artigo no *journal Multimedia Tools and Applications* em parceria com outros membros do grupo de pesquisa na UFF.

5 Conclusão

Este artigo apresentou um resumo das contribuições e dos resultados principais obtidos na tese de doutorado defendida no Programa de Pós-graduação em Computação da Universidade Federal Fluminense em setembro de 2021. O trabalho desenvolvido na tese está relacionado ao tema de aplicações multimídia e mulsemídia, e teve como objetivos o gerenciamento de sincronismo de aplicações multimídia e mulsemídia, e também a extensão da linguagem NCL e do middleware Ginga para dar suporte à especificação de efeitos sensoriais como entidades de primeira classe em sistemas de TV digital.

Os testes realizados demonstraram que a preparação automática pode auxiliar na redução do atraso de apresentação de mídia ou efeitos em aplicações mulsemídia. Além disso, o evento de preparação proposto permite que o autor da aplicação especifique o disparo de uma preparação quando achar necessário. Ou então, que o autor verifique condições de apresentação, onde o conteúdo só será apresentado se o mesmo já estiver preparado. Este novo evento possibilita o surgimento de novas aplicações, como por exemplo, entrega de propagandas direcionadas e conteúdo personalizado no ambiente de TV Digital.

Referências

- [1] ABNT. 2021. *Televisão digital terrestre - Codificação de dados e especificações de transmissão para radiodifusão digital - Parte 2: Ginga-NCL para receptores fixos e móveis-Linguagem de aplicação XML para codificação de aplicações*.
- [2] G. Ghinea et al. 2014. Mulsemimedia: State of the art, perspectives, and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 11 (2014), 17.
- [3] Z. Yuan et al. 2015. Perceived synchronization of mulsemimedia services. *IEEE Transactions on Multimedia* 17, 7 (2015), 957-966.
- [4] ISO/IEC 23005-3 2011. Information technology – Media context and control – Part 3: Sensory information.
- [5] M. Josue, R. Abreu, F. Barreto, D. P. Mattos, G. Amorim, J. A. F. Dos Santos, and D. C. Muchaluat-Saade. 2018. Modeling Sensory Effects as First-Class Entities in Multimedia Applications. In *ACM Multimedia Systems Conference, 2018, Amsterdam. MMSys 2018*. ACM, 1-12.
- [6] Marina I. P. Josué. 2021. *Preparação de Objetos de Mídia e Efeitos Sensoriais para Formatação de Documentos Mulsemídia*. Ph.D. Dissertation. Programa de Pós-graduação em Computação da Universidade Federal Fluminense.
- [7] L. F. G. Soares and R. F. Rodrigues. 2005. *Nested Context Model 3.0 Part 1 - NCM Core*. Technical Report. Informatics Department, PUC-Rio, Rio de Janeiro.
- [8] L. F. G. Soares, R. F. Rodrigues, and D. Muchaluat-Saade. 2000. Modeling, authoring and formatting hypermedia documents in the HyperProp system. *Multimedia Systems* 8, 2 (2000), 118-134.