

1 Introdução

Um documento hipermídia é formado por um conjunto de informações relacionadas. Dessa forma, além do seu conteúdo (vídeo, áudio, texto, imagem etc.), um documento hipermídia agrega uma estrutura formada pelos relacionamentos definidos pelo autor, através de algum formato, estrutura esta tão importante quanto o próprio conteúdo das informações relacionadas.

O modelo de organização hipermídia não é uma proposição recente (Bush, 1945), entretanto, foi a partir do surgimento da WWW (*World Wide Web*) (Berners-Lee et al., 1994), no início dos anos 90, que esse paradigma para estruturação das informações tornou-se popular.

Atualmente, certos cenários, como os sistemas de TV digital interativa, têm despertado inúmeros desafios no desenvolvimento de sistemas hipermídia, motivando novos trabalhos de pesquisa diretamente relacionados a tecnologias aplicadas a modelos hipermídia, como o padrão MPEG-4 (Koenen, 2002), abordado nesta dissertação.

Este capítulo descreve, de forma sucinta, as características do MPEG-4 aplicadas no contexto de sistemas hipermídia, especificando os principais ambientes que compõem a arquitetura desses sistemas. Após essa breve contextualização, o capítulo descreve os objetivos desta dissertação e detalha a estrutura deste documento.

1.1. Motivação

1.1.1. O Formato MPEG-4

O MPEG (*Moving Picture Experts Group*) define uma família de padrões para codificação digital de informações de áudio e vídeo. Seus padrões iniciais, MPEG-1 (Chiariglione, 1996) e MPEG-2 (Chiariglione, 2000), focam principalmente em aspectos relacionados à codificação e decodificação dessas

mídias. Por outro lado, a codificação do MPEG-4, diferente da codificação linear de áudio e vídeo do MPEG 1 e 2, é baseada em objetos, isto é, as cenas audiovisuais são definidas em termos de objetos de mídia (vídeo, áudio, texto, imagem etc.) codificados separadamente.

A codificação audiovisual baseada em objetos, proposta pelo MPEG-4, possui algumas vantagens: a abordagem orientada a objeto permite aos autores reusarem material audiovisual; objetos podem ser codificados usando diferentes resoluções espaciais e temporais; técnicas de compactação e compressão podem ser aplicadas individualmente, de acordo com as características de cada mídia; e objetos sintéticos, que variam desde objetos simples de duas dimensões, como linhas e pontos, até objetos complexos de três dimensões, como animações faciais e corporais, podem fazer parte de uma apresentação, em conjunto com os demais objetos de mídia naturais¹.

Isoladamente, os objetos de mídia, resultantes da codificação do conteúdo audiovisual, não definem uma apresentação. Dado um conjunto qualquer de objetos de mídia, várias apresentações distintas podem ser definidas. Portanto, para que as vantagens da abordagem orientada a objeto sejam alcançadas na codificação do conteúdo audiovisual, faz-se necessário especificar informações complementares relativas à apresentação, como a forma de associação dos objetos e suas propriedades para exibição. Faz-se também necessário realizar a entrega síncrona dessas especificações, em conjunto com os objetos de mídia codificados, aos exibidores (receptores).

No MPEG-4 as informações relativas à apresentação, tais como os relacionamentos espaço-temporais, que definem a distribuição dos objetos de mídia em relação aos dispositivos de exibição e ao tempo da apresentação, e os relacionamentos interativos, que são aqueles associados a eventos acionados pela interação do usuário, encontram-se especificadas em documentos independentes da codificação do conteúdo.

Embora a especificação das informações relativas à apresentação seja independente da codificação do conteúdo dos objetos, no MPEG-4 elas são

¹ Objetos naturais são definidos como aqueles capturados por dispositivos tais como: câmeras, microfones etc; já os objetos sintéticos são produzidos por unidades computacionais (Koenen, 2002).

multiplexadas em conjunto. As informações relativas à apresentação são codificadas em um formato binário denominado BIFS (*Binary Format for Scenes*) (ISO/IEC, 2001). Esse formato favorece o armazenamento integrado da especificação dos relacionamentos com os objetos de mídia. O conteúdo de cada objeto, por sua vez, é individualmente codificado em um formato binário, de acordo com algoritmos específicos adotados na sua codificação (ISO/IEC, 2001). Na decodificação, o fluxo multiplexado permite aos exibidores realizar simultaneamente a decodificação do conteúdo e da estrutura, obtendo uma apresentação imediata, ainda que o documento multimídia/hipermídia seja parcialmente recebido.

Segundo Koenen (Koenen, 2002) os conceitos do MPEG-4, planejados como metas na definição do padrão, podem ser sintetizados em quatro itens:

- Representação através de objetos de mídia do conteúdo audiovisual de natureza natural ou sintética;
- Multiplexação e sincronização dos conteúdos dos objetos de mídia para transporte em redes que ofereçam garantias dos requisitos de qualidade de serviço (QoS) apropriados à natureza específica do conteúdo de cada objeto;
- Possibilidade de composição dos objetos de mídia, criando objetos compostos até o nível de formação de uma cena audiovisual;
- Interação no exibidor final (exibidor da apresentação) com a cena audiovisual gerada a partir da composição dos objetos de mídia.

1.1.2. MPEG-4 e os Sistemas Hipermídia

Os conceitos do padrão MPEG-4 podem ser abordados no contexto de sistemas hipermídia, que definem uma arquitetura normalmente formada por três ambientes: o ambiente de autoria; de armazenamento; e de execução de documentos. A esses ambientes correspondem as diferentes fases usualmente envolvidas no processo de produção de documentos multimídia/hipermídia. A Figura 1.1 ilustra a definição de um sistema hipermídia completo, onde os ambientes estão localizados em estações distintas. Num sistema hipermídia distribuído, como na Figura 1.1, a troca de informações entre seus ambientes tem por suporte um provedor de serviços de comunicações.

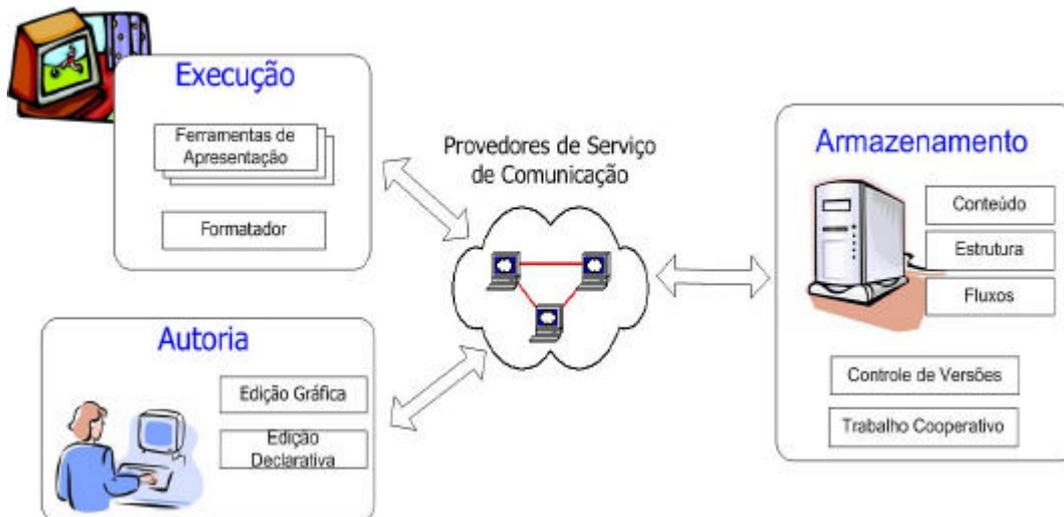


Figura 1.1 – Visão geral de um sistema hipermídia

O ambiente de autoria é o responsável por capturar as expressões do autor e defini-las através dos componentes do documento, das propriedades desses componentes, dos relacionamentos entre esses componentes e dos relacionamentos com componentes pertencentes a outros documentos. Para dar suporte às tarefas desse ambiente, ferramentas para autoria gráfica (edição gráfica) e textual (edição declarativa) podem ser utilizadas (Costa, 1996; Pinto, 2000; Moura, 2001; Coelho, 2004).

No MPEG-4, conforme anteriormente mencionado, os relacionamentos entre os objetos de mídia são especificados em um formato binário. Esse formato favorece o armazenamento e a distribuição linear da especificação das relações entre os objetos em conjunto com o seu conteúdo, no entanto, ele é inapropriado para autoria de aplicações multimídia/hipermídia.

Para facilitar a autoria, o MPEG-4 oferece duas linguagens declarativas: XMT-A e XMT-O (ISO/IEC, 2001). Em XMT-A são definidas todas as expressões existentes no formato binário, sem perda de representatividade. No entanto, ao contemplar todas essas expressões, essa linguagem tornou-se excessivamente complexa e extensa para a autoria. Como consequência, a linguagem XMT-O foi desenvolvida, com base na linguagem SMIL 2.0 (W3C, 2005a; Bulterman & Rutledge, 2004).

A linguagem XMT-O utiliza composições, que podem conter objetos de mídia e outras composições, para estruturar um documento. Nessa linguagem, as composições possuem semântica de sincronização, o que torna, a princípio, mais fácil a tarefa de autoria. No entanto, esse benefício é limitado pela existência de

um conjunto simples de composições (paralela, seqüencial e exclusiva) (ISO/IEC, 2001), o que dificulta a definição de relacionamentos complexos, podendo ser necessário, para esses casos, estabelecer composições com vários níveis de aninhamento. Além disso, essas composições obrigam o autor a estruturar o documento de acordo com a especificação para apresentação (Muchaluat-Saade, 2003).

No contexto da autoria, esta dissertação tem como foco facilitar a concepção pelos autores de documentos MPEG-4. Ao adotar a linguagem XMT-O para essa tarefa, esta dissertação propõe a utilização de templates hipermídia voltados para as características dessa linguagem, permitindo a definição de composições com outras semânticas, além das já existentes, favorecendo o reuso e, conseqüentemente, o processo de autoria.

Conceitualmente, templates hipermídia podem especificar diversos tipos de relacionamentos entre os objetos, se constituindo em estruturas hipermídia genéricas, que podem ser herdadas por composições. Como caso particular, na linguagem XMT-O, os templates podem definir relações de inclusão recursivas, oferecendo semânticas temporais a uma única composição, que antes somente eram obtidas através de composições com vários níveis de aninhamento.

Os templates adotados foram definidos inicialmente no módulo *XTemplate* (Muchaluat-Saade, 2003) de NCL 2.0 (*Nested Context Language*) (Silva et al., 2004b). NCL é uma linguagem declarativa e modular XML (W3C, 2000), baseada no modelo NCM (*Nested Context Model*) (Soares et al., 2003), que introduziu uma série de novos conceitos. Posteriormente, adaptações a esse módulo tornaram-no mais flexível, reestruturando suas funcionalidades em diferentes perfis, direcionados às linguagens de autoria onde ele pode ser empregado. Particularmente, o perfil SMIL+*XTemplate* (XT+SMIL) (Silva et al., 2004a), voltado para a linguagem SMIL 2.0, foi adotado neste trabalho como modelo base, uma vez que a linguagem XMT-O baseia-se em SMIL 2.0, possuindo estruturas semelhantes.

Documentos multimídia/hipermídia têm, geralmente, os objetos de mídia (vídeo, áudio, texto, imagem etc.) que os compõem e a especificação dos seus relacionamentos, armazenados em separado. Essa separação é proposta, principalmente, pelos requisitos exigidos pelos diferentes tipos de mídia, particularmente mídias contínuas (áudio e vídeo). Para essas mídias, quando o

armazenamento é realizado de forma distribuída, é desejável que o serviço de entrega dos dados seja capaz de estabelecer e manter acordos de qualidade de serviço (QoS) intramídia (Rodrigues, 2003).

No MPEG-4 o armazenamento dos documentos multimídia/hipermídia pode, eventualmente, ser realizado de forma a separar a especificação dos relacionamentos dos objetos relacionados. No entanto, ao contrário do citado no parágrafo anterior, no MPEG-4 pode-se considerar que o objetivo dessa separação consiste em facilitar as atualizações e modificações nos documentos. Nos documentos, durante a sua concepção, ou mesmo posteriormente, atualizações e modificações são usuais, principalmente quando eles são gerados de forma colaborativa, com múltiplos autores simultâneos ou mesmo subseqüentes (Gorini, 2001).

No entanto, como já mencionado na Seção 1.1.1, a principal forma de armazenamento dos documentos multimídia/hipermídia no MPEG-4, ao contrário da proposta tradicional de separação, segue um modelo integrado, onde as relações entre os objetos de mídia e seus conteúdos permanecem encapsuladas em uma estrutura única, formada por fluxos (*streaming*) que, posteriormente, são diretamente distribuídos aos exibidores.

Cada objeto de mídia pode corresponder a um fluxo e receber o tratamento apropriado à natureza específica do seu conteúdo (codificação individual, requisitos de QoS específicos para o transporte etc). Em alguns casos, objetos podem estar codificados em mais de um fluxo. A relação entre o objeto de mídia e o número de fluxos utilizados para sua codificação é uma opção que oferece aos exibidores padrões de codificação distintos, onde cada fluxo pode corresponder a um padrão diferente ou mesmo a um único padrão, com pequenas variações. Nessa abordagem, também é possível que um mesmo objeto seja codificado de forma escalável e que seus níveis de escalabilidade sejam transmitidos em vários fluxos individuais. Por exemplo, um vídeo pode ser codificado em três níveis de qualidade; a utilização no exibidor dos três fluxos permite apresentar o vídeo em alta qualidade, a utilização de apenas dois fluxos oferece uma apresentação de qualidade intermediária e, a disponibilidade de apenas um fluxo, implica em um vídeo de baixa qualidade.

O modelo de armazenamento integrado padronizado pelo MPEG-4, além de preservar os requisitos exigidos pelos diferentes tipos de mídia, é particularmente

útil em cenários onde a sincronização das mídias, principalmente daquelas provenientes de fontes distintas, é difícil de ser oferecida, como na distribuição de conteúdo em plataformas de TV interativa. Nos sistemas de TV interativa, além do conteúdo de áudio e vídeo principais que, tradicionalmente compõem a apresentação dos usuários (clientes), existe a possibilidade de que aplicações multimídia/hipermídia sejam adicionalmente transmitidas, possibilitando novas representações para as informações, tornando-as mais atrativas e inovadoras, quando comparadas às apresentações tradicionais. A codificação do MPEG-4, aplicada a esses sistemas, permite que todo o conteúdo, incluindo o audiovisual principal e as aplicações, independente da origem ou funcionalidade, seja codificado através de objetos de mídia, integrados à especificação dos seus relacionamentos.

A especificação dos relacionamentos em formato BIFS, anteriormente mencionado, corresponde a um novo fluxo, em sincronia com os demais fluxos dos objetos de mídia que relaciona. É importante mencionar que as especificações em BIFS, relativas à sincronização dos objetos de mídia são, por princípio, temporalmente lineares, isto é, as especificações temporais de sincronismo desse formato são relativas ao tempo de início da apresentação e não aos tempos dos demais objetos, definindo uma estrutura de dados em *timeline*, formada pelos objetos e seus respectivos instantes de exibição. A estrutura em *timeline* não permite determinar as especificações originais concebidas pelos autores, premissa básica necessária para a adoção de um formatador hipermídia (Rodrigues, 2003).

No contexto do armazenamento, o foco deste trabalho está inserido no desenvolvimento de ferramentas para a conversão de documentos multimídia/hipermídia, especificados através da linguagem NCL 2.0, para as linguagens do padrão MPEG-4. Essa conversão permite, entre outros benefícios, que documentos em NCL 2.0 possam ser armazenados e distribuídos de forma integrada, como documentos MPEG-4.

Nessa conversão, documentos especificados na linguagem NCL 2.0 podem ser transformados para a linguagem XMT-O, escolhida por ter uma representatividade mais próxima de NCL 2.0, em relação às outras linguagens e formatos do MPEG-4. Os documentos obtidos em XMT-O, a partir da conversão das especificações em NCL 2.0, podem ser convertidos, tanto para linguagem XMT-A, quanto para o formato binário. É no caso do formato binário, que a

multiplexação da especificação dos relacionamentos com os conteúdos dos objetos de mídia gera um formato para armazenamento (ou transmissão) integrado. Para complementar o intercâmbio entre os formatos, um conversor da linguagem XMT-O para NCL 2.0 também foi desenvolvido, possibilitando a exibição de programas MPEG-4 em formatadores NCL (Rodrigues, 2003).

Cabe ressaltar que a conversão entre os formatos agrega outros benefícios como, por exemplo, permitir que ferramentas de autoria e exibição desenvolvidas para um formato possam ser utilizadas por outro. Como caso particular, o ambiente de autoria NCL poderá ser usado na concepção de documentos MPEG-4 e documentos NCL poderão ser apresentados em exibidores MPEG-4 e vice-versa.

Como resultado da concepção de documentos multimídia/hipermídia, tem-se, primordialmente, apresentações audiovisuais interativas que devem ser exibidas aos usuários no ambiente de execução. É nesse ambiente que as especificações são interpretadas e apresentadas, por meio de dispositivos de saída correspondentes a alguma plataforma particular de exibição. Nesse ambiente também estão disponíveis os mecanismos de interação com os usuários, através de dispositivos de entrada, restringindo, no entanto, as possibilidades de interação àquelas definidas previamente pelos autores na especificação dos documentos.

Em uma visão geral (Figura 1.1), o formatador é o responsável pela exibição de um documento, possuindo um conjunto de ferramentas de exibição para o controle das características dos diversos objetos de mídia e por suas associações com os dispositivos de entrada e saída. Cada ferramenta de exibição é desenvolvida para um conjunto de objetos de mídia e gerencia aspectos relacionados à descompressão, decodificação, controle da taxa de apresentação, interação, entre outros (Rodrigues, 2003).

Compondo o formatador, um orquestrador intermídia é o responsável pelo controle da apresentação, preservando os relacionamentos definidos pelos autores. O orquestrador interage diretamente com as ferramentas de exibição que repassam, para cada objeto, o estado da apresentação, bem como as interações realizadas pelos usuários. Além de controlar as especificações originais dos autores, cabe ao formatador manter a consistência da apresentação através de ajustes no documento hipermídia. Esses ajustes podem ser necessários devido a ações realizadas pelos usuários, gerando eventos não previstos, devido a atrasos

no sistema de comunicação entre o ambiente de execução e o de armazenamento etc. (Rodrigues, 2003).

As ferramentas de exibição MPEG-4 são desenvolvidas para apresentação de cenas audiovisuais, onde cada cena é definida como uma composição que pode conter vários objetos, bem como os relacionamentos entre eles. Dentro de cada cena audiovisual não existem restrições para a definição de relacionamentos entre os objetos definidos internamente, no entanto, os relacionamentos entre cenas distintas são limitados. Nesse caso, o relacionamento, normalmente, possui semântica de referência, isto é, a cena de origem, quando apresentada em um exibidor MPEG-4, possui uma âncora associada a um objeto que, ao ser acionada, interrompe o desenvolvimento dessa cena e carrega a cena de destino, relacionada através de um elo.

No ambiente de execução, o foco deste trabalho está na integração de uma ferramenta de exibição MPEG-4 ao formatador HyperProp, que já possui implementadas ferramentas de exibição para tratar os principais tipos de mídia (texto, imagens, áudio e vídeo). Esta integração permitirá não apenas a exibição de objetos MPEG-4 em formatadores NCL, mas também permitirá a definição de novos relacionamentos entre cenas MPEG-4, superando a limitação atual, ou mesmo entre cenas MPEG-4 e composições, ou objetos, de outros tipos.

1.2. Objetivos

Esta dissertação tem como principal objetivo prover a integração dos recursos oferecidos pelo MPEG-4, contextualizados nos diferentes ambientes que compõem um sistema hipermídia completo, às características da linguagem NCL.

A proposta de integração é composta pelo desenvolvimento dos conversores entre a linguagem NCL 2.0 e XMT-O e vice-versa, além de integrar conversores dos documentos em XMT-O, obtidos através da conversão de NCL 2.0, para XMT-A e BIFS. Nos conversores entre as linguagens declarativas, um *meta-framework* para linguagens XML modulares é utilizado. Para o formato binário, são utilizados os softwares de referência do MPEG-4, e também implementações baseadas nesses softwares.

Complementarmente, no ambiente de execução, este trabalho integra uma ferramenta de exibição MPEG-4 ao sistema HyperProp, para que documentos especificados na linguagem NCL possam estabelecer relacionamentos entre cenas MPEG-4, ou mesmo entre cenas MPEG-4 e outras composições, ou mesmo objetos, de tipos diversos. Além de integrar as ferramentas existentes, este trabalho também pretende estendê-las com o objetivo de tratar as cenas MPEG-4 como composições efetivamente e não apenas como objetos, conforme apresentado no Capítulo 4.

Outro objetivo, desta vez diretamente relacionado à autoria de documentos MPEG-4, é a proposição de templates para composições da linguagem XMT-O, cuja função é facilitar a autoria através do reuso de estruturas de composições complexas. Os templates devem ser derivados daqueles propostos para SMIL 2.0, apresentados no Capítulo 5.

A proposta de integração do MPEG-4, em toda a sua amplitude, entre outras possibilidades, permite se pensar na alternativa de uso do formatador HyperProp como middleware do terminal de acesso para o Sistema Brasileiro de TV Digital, bem como a NCL como a linguagem para desenvolvimento de aplicações interativas, sem a perda de interoperabilidade com outros padrões, ainda que futuros, como no caso, o MPEG-4.

1.3. Organização da Dissertação

Esta dissertação encontra-se organizada como a seguir. O Capítulo 2 apresenta os conceitos do padrão MPEG-4 em detalhes, e os conceitos da linguagem NCL necessários ao entendimento do trabalho. Após descrever a arquitetura do padrão MPEG-4, o capítulo concentra-se nas suas linguagens de autoria, detalhando suas estruturas sintáticas, que serão necessárias nas conversões, e também em seus aspectos semânticos incluindo-se, principalmente, as peculiaridades dos relacionamentos entre os objetos de mídia em cada uma das suas linguagens. Por fim, o capítulo apresenta os conceitos básicos do modelo NCM e da sua representação declarativa (a linguagem NCL), com foco nos templates para autoria de documentos hipermídia.

O Capítulo 3 apresenta os conversores desenvolvidos entre as linguagens NCL 2.0 e XMT-O, e também o conversor dos documentos em XMT-O, obtidos através da conversão de NCL 2.0, para XMT-A, também desenvolvido nesta dissertação, e BIFS.

O Capítulo 4 trata da integração de ferramentas de exibição MPEG-4 ao formatador HyperProp. O capítulo detalha as possibilidades de relacionamentos entre cenas MPEG-4 e entre cenas e outras composições, ou objetos distintos. Nesse capítulo é proposto que cenas audiovisuais MPEG-4 sejam tratadas como composições, apresentando uma ferramenta de exibição estendida para esse fim.

O Capítulo 5 aborda a extensão dos perfis de templates, propostos inicialmente para a linguagem NCL. Ênfase especial é dada ao perfil XT+SMIL, base dos templates apresentados nesta dissertação, também abordados no Capítulo 5 com a denominação de perfil XT+XMT-O.

Finalmente, o Capítulo 6 tece as considerações finais desta dissertação, destacando as suas contribuições e os trabalhos futuros.