

# Autoria de Aplicações Multissensoriais para TV 3.0 com a Ferramenta STEVE

Rômulo Vieira  
romulo\_vieira@midia.com.uff.br  
Laboratório MídiaCom  
Universidade Federal Fluminense  
Niterói, Rio de Janeiro, Brasil

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

Raphael Abreu  
raphael.abreu@midia.com.uff.br  
Laboratório MídiaCom  
Universidade Federal Fluminense  
Niterói, Rio de Janeiro, Brasil

Joel A. F. dos Santos  
jsantos@eic.cefet-rj.br  
GPMM  
Centro Federal de Educação  
Tecnológica Celso Suckow da Fonseca  
Rio de Janeiro, Rio de Janeiro, Brasil

Douglas Mattos  
douglas@midia.com.uff.br  
Laboratório MídiaCom  
Universidade Federal Fluminense  
Niterói, Rio de Janeiro, Brasil

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

## Resumo

Os avanços no setor televisivo brasileiro sempre estiveram atrelados com as inovações tecnológicas de cada época. Seja no seu início que suportava apenas transmissões ao vivo, passando pela inserção de conteúdo gravado em fitas magnéticas até a adoção de cores, as transmissões sempre foram marcadas pela busca por uma maior qualidade e interatividade. Atualmente, o Fórum do Sistema Brasileiro de Televisão Digital Terrestre (SBTVD) está trabalhando em um novo padrão nacional, chamado de TV 3.0, que dentre várias diretrizes, busca incorporar aspectos multissensoriais. Entretanto, pessoas sem noções de programação ou com pouco domínio da tecnologia encontram severas dificuldades na criação deste tipo de conteúdo. Para preencher esta lacuna, o presente artigo apresenta a ferramenta STEVE, uma ferramenta de autoria baseada em linha temporal que tem exatamente este tipo de público como alvo. Para mais, é apresentado um caso de uso para validar esta ferramenta, assim como um detalhamento de como ela pode ser usada para geração automática de aplicações NCL 4.0, com total aplicabilidade no vindouro sistema de TV no Brasil.

**Keywords:** TV 3.0, mulsemídia, STEVE, autoria, efeitos sensoriais, NCL 4.0

## 1 Introdução

O marco inicial da TV no Brasil se dá no ano de 1950, quando ocorreu a primeira transmissão desse tipo no país, inaugurando também aquela que ficaria posteriormente conhecida

como “TV 1.0”. Como a única forma de registro de sons e imagens neste período era através de películas cinematográficas, toda a programação era ao vivo [3, 19].

A partir desse ponto, diversas transformações tecnológicas tiveram início. Isso incluiu avanços como a capacidade de capturar áudio diretamente na película, bem como a incorporação da tarja magnética e do registro sonoro óptico. Esses avanços viabilizaram a gravação de conteúdo em fitas magnéticas (também conhecidas como *video tapes*) e sua subsequente edição, elevando o nível de complexidade dos programas televisivos. Como resultado, apenas certos gêneros televisivos, como telejornais, eventos esportivos e programas de auditório com interação do público, foram mantidos no formato de transmissão ao vivo [3].

A partir da década de 1970, o sistema brasileiro, que até então adotava imagens em preto e branco, com 525 linhas por quadro e 30 quadros por segundo (conhecido como padrão M), passou a adotar o padrão de cores europeu, com 625 linhas, dando origem ao sistema PAL-M, considerado um dos melhores modelos de cor do mundo devido a sua estabilidade e compatibilidade com outras normas. Na esteira disso, novas tecnologias foram inseridas nos anos subsequentes, como o início da distribuição por satélites em substituição ao método terrestre de micro-ondas, o uso de antenas parabólicas e a implantação do som estéreo, que empregava dois canais monoaurais independentes e sincronizados no tempo. Esse avanço visava aprimorar a experiência auditiva dos espectadores e criar um ambiente de áudio mais realista, inaugurando também a “TV 1.5” [28].

Em 2006, após um decreto presidencial [9], foi criado o Fórum do Sistema Brasileiro de Televisão Digital Terrestre (SBTVD-T), composto por membros do governo, da academia, da indústria e das emissoras, com o objetivo de especificar as políticas e técnicas para implementação da televisão digital terrestre no país. A primeira transmissão nestes moldes se deu no ano seguinte, marcando também o começo da fase “TV 2.0” [4].

---

In: Workshop Futuro da TV Digital Interativa, Ribeirão Preto, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2023.

A etapa seguinte, denominada “TV 2.5”, passou a explorar o desenvolvimento de duas novas funcionalidades. A primeira delas era focada na integração *broadcast-broadband*, envolvendo a criação de um novo perfil de receptor para o *middleware* Ginga (denominado perfil D ou DTV Play), permitindo vídeo sob demanda, sincronização de várias telas e entrega de conteúdo direcionado. A segunda aplicação visava melhorar o conteúdo audiovisual a partir de três novos codificadores de áudio imersivos, a saber: MPEG-H, E-AC 3 JOC e AC-4, e vídeo HDR (*High Dynamic Range*) nos formatos SL-HDR1 e HLG. Importante ressaltar que esta é a atual geração de TV Digital terrestre no país [30].

A partir do aprimoramento dos aparelhos televisores e da integração contínua deles com a Internet, novas funcionalidades foram surgindo, desde aprimoramento na imagem e som, até a eclosão de conteúdos 3D e recursos multimídia. Estas mudanças representam uma quebra no paradigma inicial da TV, baseada em transmissões unilaterais onde somente as emissoras eram responsáveis pelo envio de conteúdo, permitindo que este equipamento seja uma ferramenta de emissão do público, isto é, tornando-o capaz de também configurar o conteúdo, modelando-o de acordo com seus filtros e preferências.

Todas estas mudanças e ensejos levaram o Fórum SBTVD a iniciar uma transição para a fase “TV 3.0” [18], que visa contemplar diversos aspectos, como navegação baseada em aplicativos, ruptura da linearidade da grade de programação, integração com Internet das Coisas (IoT), *streaming* adaptativo e aumento da imersão via realidade virtual e efeitos sensoriais, para citar alguns.

No que diz respeito a este último aspecto, ele já vem sendo explorado pela comunidade científica através de um conceito chamado mulsemídia (*mulsemedia - Multiple Sensorial Media*) [20]. Este campo oferece aos usuários novas formas de experiência a partir do uso dos cinco sentidos humanos (audição, visão, olfato, paladar e tato) em aplicações multimídia interativas.

Para contemplar estes aspectos no contexto de TV digital, foi desenvolvida a versão 4.0 para a linguagem NCL (*Nested Context Language*). Ela permite a sincronização entre efeitos sensoriais e mídias tradicionais, além de propor operações para preparação desses conteúdos. Por conta destas funcionalidades, NCL 4.0 foi adotada como tecnologia para codificação de aplicações multissensoriais no Projeto TV 3.0.

Entretanto, verifica-se uma falta de ferramentas de autoria para este tipo de conteúdo, em especial aquelas baseadas em interfaces gráficas e voltadas para um público sem pleno conhecimento de programação. Os ambientes de autoria serão fundamentais para adoção da tecnologia sendo proposta para o novo padrão de TV.

Neste contexto, o presente artigo exhibe uma ferramenta nesses moldes, chamada STEVE, apresentando sua estrutura, modo de funcionamento e alguns casos de uso que corroboram para sua adoção na TV 3.0. Maiores detalhes sobre

sua geração automática de documentos NCL 4.0 também são fornecidos.

O restante do artigo está disposto da seguinte maneira. A Seção 2 aborda a fundamentação teórica sobre autoria mulsemídia. A Seção 3 apresenta o editor gráfico STEVE. A Seção 4 discute a funcionalidade de extração automática de efeitos sensoriais sincronizados com o conteúdo audiovisual no STEVE. A Seção 5 discute a geração automática de código NCL 4.0 pelo STEVE seguindo a padronização a ser adotada na TV 3.0. A Seção 6 apresenta as limitações da ferramenta STEVE para autoria NCL 4.0. Finalmente a Seção 7 apresenta as considerações finais e trabalhos futuros.

## 2 Autoria Mulsemídia

### 2.1 Modelos e Linguagens Mulsemídia

Normalmente, aplicações mulsemídia são definidas especificadas em documentos, um tipo de arquivo que estrutura as mídias interativas com os efeitos sensoriais, sincronizando-as a partir do tempo ou a partir da ocorrência de eventos. Eles são baseados em modelos conceituais representados por nós, elos (*links*) e elementos de composição. A relação temporal entre eles, por sua vez, pode ser especificada a partir de diferentes paradigmas [12, 17].

Um deles é o modelo baseado em *scripts*, que utiliza linguagens de programação imperativas para definir o comportamento da apresentação do documento. Outro paradigma é o modelo baseado em linha do tempo, aconselhável quando o autor da aplicação sabe, *a priori*, a duração dos objetos ou quando consegue manipulá-los de forma explícita. Uma outra opção é a representação baseada em grafos que pode ser dividida em dois grupos: fluxogramas, que descrevem o estado da aplicação ao longo de sua execução, e grafos direcionados [6], como redes de Petri [27] ou diagrama de estados [15].

Há ainda o paradigma baseado em estruturas [5, 7, 33], no qual composições possuem uma semântica temporal embutida. Dois tipos comuns são as composições sequenciais (elementos filhos apresentados de forma sequencial) e paralelas (elementos filhos apresentados simultaneamente). Outro paradigma é o baseado em restrições, nos qual o autor especifica regras que, quando avaliadas, determinam os limites temporais para sincronizar entre os nós do documento.

No que diz respeito à sincronização baseada em pontos de referência, esta abordagem utiliza marcos específicos (tais como o início ou o fim de um evento) como referências temporais para estabelecer a sincronização entre eles. Por outro lado, o paradigma baseado em eventos se fundamenta nas ocorrências que ocorrem durante a execução de uma aplicação, incluindo eventos de apresentação, seleção, atribuição e preparação, ao passo que os eventos síncronos se relacionam a momentos específicos, como o início, a interrupção e a pausa na apresentação de um item de mídia. O evento de seleção representa a interação do usuário. A atribuição, por

sua vez, envolve a associação de um valor ou comando a um objeto, e o evento de preparação se refere ao processo de preparar o conteúdo de um objeto de mídia para exibição ou antecipar os comandos para um atuador que vai renderizar um efeito sensorial antes da exibição propriamente dita.

A partir disso, surge uma pletera de especificações inspiradas nestes conceitos, como o padrão MPEG-V [26], que fornece um conjunto de linguagens baseadas em XML para permitir o intercâmbio de informações entre objetos do mundo real, como sensores e atuadores, e entidades do mundo virtual, no caso uma aplicação mulsemídia. Uma das linguagens definidas pelo padrão emprega uma abordagem baseada em linha do tempo para realizar sincronização temporal entre efeitos sensoriais e conteúdo audiovisual [23].

Outro formato baseado em XML é a linguagem NCL 4.0 [24], uma extensão de NCL 3.0 [32] na qual efeitos sensoriais são tratados como entidades de primeira classe [23]. Tal qual em um tipo de mídia tradicional, os autores definem propriedades para os elementos de efeito e especificam a posição dos efeitos em coordenadas esféricas, tornando-os independentes da instalação física na qual a aplicação é executada e possibilitando sua integração com mídias renderizadas em ambientes 360 graus. NCL 4.0 também facilita a interação multimodal e oferece suporte a vários usuários. NCL 4.0 foi selecionada como uma das tecnologias a ser utilizada no Projeto TV 3.0 para a camada de codificação de aplicações.

## 2.2 Ferramentas Mulsemídia

No contexto de aplicações que exploram o conceito de *crowdsourcing*, destaca-se o CrowdMuse [21], que atua como uma ferramenta para detectar a presença de efeitos sensoriais em um conteúdo audiovisual com base no público e retransmiti-los aos autores do serviço. No entanto, sua ênfase é principalmente na edição de vídeo, em vez de fornecer um ambiente de autoria abrangente. O H-Studio [10], por sua vez, concentra-se em suportar dois efeitos sensoriais específicos, vibração e movimento, fornecendo uma linha de tempo para sincronização e um menu para definir os parâmetros do efeito.

Várias ferramentas empregam o paradigma da linha do tempo para sincronizar efeitos sensoriais com conteúdo audiovisual. SEVino (*Sensory Effect Video Annotation*) [34] oferece canais representando diferentes tipos de efeitos sensoriais que podem ser sincronizados com conteúdo de vídeo e suporta exportação para o formato MPEG-V. Da mesma forma, RoSE (*Representation of Sensory Effects*) [8] fornece uma interface gráfica para definir a posição espacial de efeitos sensoriais no espaço 3D e permite a exportação para arquivos MPEG-V SEM (*Sensory Effect Metadata*). SMURF (*Sensible Media Authoring Factory*) [25] apresenta uma linha do tempo para sincronização e suporta a criação de grupos de efeitos para compor efeitos sensoriais mais complexos, enquanto o Real4DASudio [31] oferece um recurso de simulação 3D para validação de efeitos e emprega interfaces baseadas em mídia e em eventos para sincronização de efeitos.

Ainda baseado nestes conceito funcional, está o SCENAPROD [22], inspirado em *softwares* de edição de mídia, com o objetivo de permitir a criação de aplicações interativas voltadas para a área médica sem que o autor precise utilizar programação. Um diferencial dessa abordagem está na liberdade que ela oferece aos usuários para manipular a linha do tempo a partir da temporalidade real (quando ela é baseada em eventos sequenciais e cronológicos) ou a partir da temporalidade do evento (quando é baseada em ações que podem ocorrer a qualquer momento).

## 3 Editor Gráfico STEVE

O STEVE (*Spatio-Temporal View Editor*) [13] surgiu como um editor espaço-temporal para autoria de documentos multimídia interativos. Seu modo de funcionamento é inspirado no modelo baseado em eventos SIMM (*Simple Interactive Multimedia Model*), que atribui uma interpretação causal às relações temporais de Allen [2]. Desse modo, o STEVE permite que os autores sincronizem itens de mídia a partir de relações gráficas. Além disso, fornece especificações das propriedades de cada objeto e provê uma pré-visualização do conteúdo criado. O editor também permite a simulação de eventos de interação dos usuários, fornece *feedback* aos autores em relação à consistência temporal, e exporta o conteúdo multimídia para as linguagens NCL e HTML5.

Entretanto, a primeira versão do STEVE permite apenas a edição de conteúdo multimídia, não abrangendo efeitos sensoriais. Para preencher esta lacuna e permitir a criação de aplicações para sistemas de TV Digital e web que sigam estes moldes, foi desenvolvido o STEVE 2.0 [14, 29]. Esta variante apresenta uma mudança significativa em sua estrutura, substituindo o modelo SIMM pelo MultiSEM (*Multimedia Sensory Effect Model*) [14, 29], um modelo de domínio específico baseado em eventos, que visa aprimorar o desenvolvimento de aplicações mulsemídia e apoiar a implementação de ambientes de autoria gráfica baseados em linhas do tempo para auxiliar autores sem habilidades de programação. Sendo assim, esta nova versão do STEVE define o comportamento temporal de suas aplicações da mesma maneira com relações temporais, refletindo diretamente na estrutura e no funcionamento de sua interface gráfica.

Ainda sobre sua interface, ela é composta por um repositório de conteúdo multimídia no canto superior esquerdo, um painel central para os usuários editarem propriedades de apresentação de mídia e características de renderização de efeitos sensoriais, um painel de visualização para reproduzir conteúdo audiovisual no canto superior direito e um painel temporal para visualizar a sincronização entre as mídias e efeitos sensoriais na região inferior. Ademais, propriedades de apresentação de conteúdo multimídia tradicional também podem ser alteradas, como aplicar transparência em um vídeo ou editar os parâmetros sonoros de um arquivo de áudio. A interface do editor é apresentada na Figura 1. Após a fase

de produção, as aplicações mulsemídia criadas no STEVE 2.0 podem ser exportadas para documentos NCL 4.0, suportados pelo middleware Ginga-NCL estendido para suportar efeitos multissensoriais [12]. A exportação em HTML5 não suporta efeitos sensoriais, apenas conteúdo de mídia tradicional.

Além disso, essa mesma interface fornece *feedback* aos autores sobre inconsistências de sincronização temporal e permite que os usuários criem relações de interatividade a partir de eventos que relacionam os objetos de mídia e os objetos de efeito. A tela para criar relação de interatividade é ilustrada na Figura 2. Adicionalmente, características gerais de renderização dos efeitos, como intensidade, tipo de aroma, frequência de luz e posições físicas de efeitos sensoriais também podem ser definidas no editor gráfico. STEVE foi implementado em linguagem Java.

#### 4 Extração automática de efeitos

STEVE 2.0 também permite a extração automática de efeitos sensoriais por meio do componente inteligente chamado STEVEML (STEVE *Machine Learning*) [1, 11]. O STEVEML segue o formato de componente inteligente apresentado em Abreu et al. [1] que propõe conjunto de diretrizes e uma implementação a ser incorporada em uma ferramenta de autoria. Com isso, a ferramenta pode incorporar algoritmos de reconhecimento de conteúdo multimídia e fornecer um mecanismo para adaptar a resposta para anotar efeitos sensoriais na linha do tempo. O componente inteligente STEVEML desenvolvido na ferramenta STEVE emprega um serviço online de reconhecimento de conteúdo de vídeo para detectar os momentos em que efeitos sensoriais poderiam ser sincronizados ao conteúdo audiovisual. Com isso, este componente sugere ao autor a inserção destes efeitos na linha do tempo do STEVE 2.0.

Para realizar esse processo na interface do editor STEVE, o autor deve clicar sobre o objeto de mídia e no menu de contexto apresentado selecionar a extração de efeitos. Uma caixa de texto é exibida para que o autor selecione quais tipos de efeitos sensoriais ele quer reconhecer no objeto de mídia escolhido. Depois da seleção, o processo de reconhecimento é iniciado e os efeitos sensoriais são adicionados na visão temporal de acordo com os tempos reconhecidos pelo componente STEVEML. O autor pode então fazer um ajuste fino modificando manualmente os efeitos na interface do STEVE 2.0.

A implementação do STEVEML segue uma arquitetura baseada na comunicação com *softwares* de reconhecimento e, na tradução do resultado do reconhecimento para efeitos sensoriais. Na implementação do STEVEML, foi usada a API para reconhecimento de vídeo<sup>1</sup>. Esta API provê reconhecimento de elementos presentes em imagens e vídeos. No caso de ser utilizado um vídeo, a API retorna um conjunto de *tags*

que indicam os elementos presentes em cada trecho de 1 segundo do vídeo.

De posse desse conjunto de *tags*, o componente STEVEML define os momentos de ativação e desativação de efeitos sensoriais. Percorrendo as *tags* presentes para cada segundo do vídeo, STEVEML define como momento de ativação de um efeito, o trecho em que a *tag* relacionada a este efeito apareceu pela primeira vez. Este efeito deverá ficar ativo enquanto a *tag* continuar aparecendo nos trechos subsequentes. O momento de desativação é definido quando a *tag* deixa de aparecer.

O processo acima depende da identificação de qual (ou quais) *tag* está relacionada a um determinado efeito sensorial. Esta relação depende do *software* de reconhecimento utilizado, dado que cada *software* diferente define diferentes *tags* e das preferencias do autor. Por exemplo, um autor pode preferir associar uma *tag* "rain" a um efeito de *spray* d'água, enquanto outro pode preferir a relação com um efeito de cheiro. Para permitir a customização da relação entre *tags* e efeitos, STEVEML provê um dicionário no formato JSON (*JavaScript Object Notation*), como pode ser visto na Listagem 1.

```

1 {
2   WIND:{" air ", " storm ", " flight " },
3   WATER_SPRAYER:{" rain ", " waterfall " },
4   VIBRATION:{" gunshot ", " explosion ", " crash
5     " },
6   TEMPERATURE:{" heat ", " cold ", " sun ", " snow "
7     , " summer ", " winter " },
8   SCENT:{" flower ", " garden ", " forest ", "
9     trees " },
10  LIGHT:{" sun ", " bright ", " light " },
11  FOG:{" fog ", " clouds ", " smoke " },
12  FLASH_LIGHT: {" lightning ", " gunshot ", "
13    thunder ", " explosion " },
14  RAINSTORM:{" storm " }
15 }
```

**Listagem 1.** Arquivo de mapeamento de *tags* em efeitos sensoriais.

Para customizar a relação entre *tags* e efeitos sensoriais o autor deve editar o arquivo de dicionário e inserir as *tags* relacionadas a cada efeito sensorial da lista. Para incluir mais *tags*, o autor deve saber quais são os possíveis retornos da rede neural. No exemplo da Listagem 1, as *tags* estão em inglês, pois este é o padrão de retorno da rede neural utilizada.

Um vídeo ilustrando a extração automática de efeitos sensoriais no STEVE está disponível neste link<sup>2</sup>.

<sup>1</sup><https://clarifai.com>

<sup>2</sup><https://youtu.be/0OziKkuMeVQ>

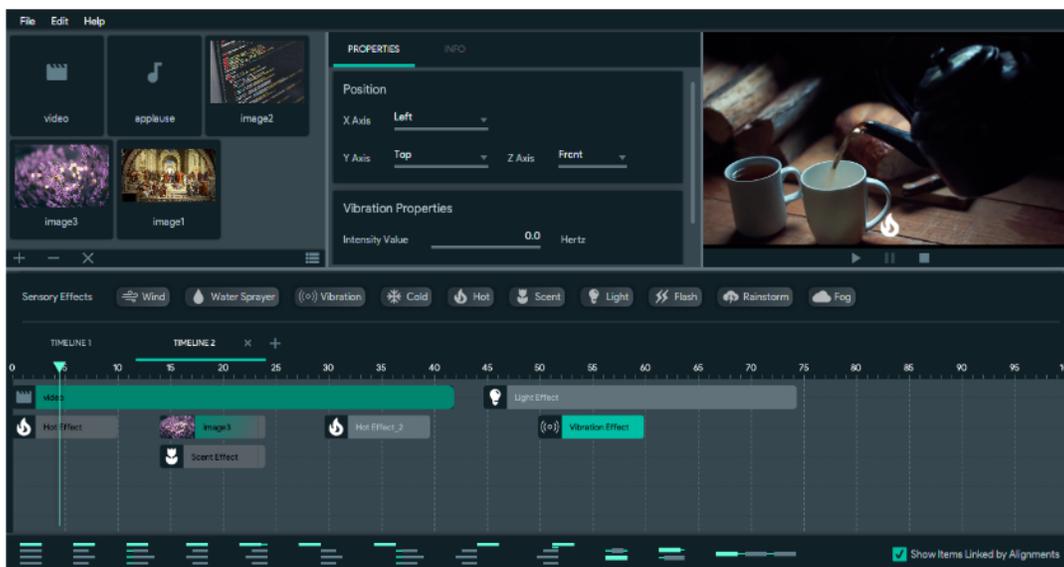


Figura 1. Interface gráfica do STEVE 2.0 [12].

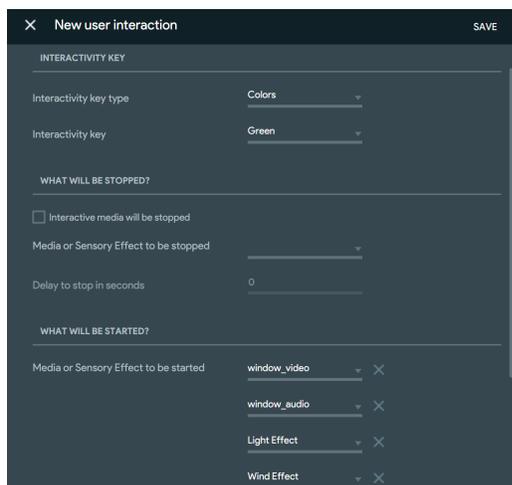


Figura 2. Interface de configuração de interatividade e de efeitos sensoriais.

## 5 Exportação Automática para NCL 4.0

### 5.1 Estudo de caso

Para validar a utilização do STEVE para autoria de aplicações mulsemídia, conforme espera-se que ocorra na TV 3.0, foi criado um caso de uso de uma aplicação imersiva com múltiplos efeitos sensoriais que auxiliam na realização de atividades cognitivas para o público idoso.

Para esta prática, é esperado que os usuários estejam sentados em uma poltrona no centro de uma sala que contém atuadores e sensores espalhados pelo ambiente. No início, a tela exibe uma representação dessa sala a partir de um nó de mídia contendo uma imagem (denominada background\_image).

acendendo com um áudio respectivo a essa ação. Ao mesmo tempo, um efeito de luz também é disparado para representar a luminosidade das chamas. Passados outros 5 segundos, um efeito de temperatura é disparado para aquecer o ambiente.

Caso o usuário interaja com o conteúdo pressionando a tecla verde de seu controle remoto, é iniciada uma animação das cortinas e janelas se abrindo em paralelo com efeitos de fluxo de vento e mudanças nas propriedades da luz e temperatura, que agora é reduzida. Um desenho esquemático deste ambiente multissensorial é observado na Figura 3.

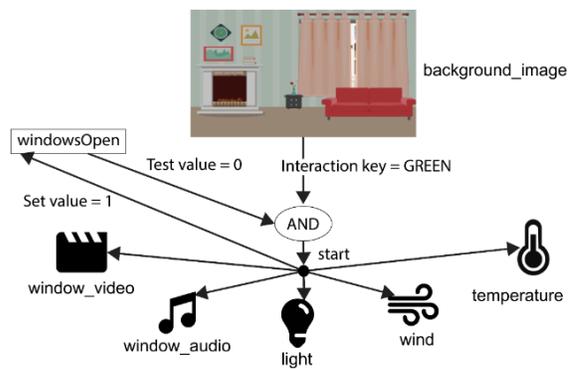


Figura 3. Estrutura do ambiente implementado como estudo de caso.

Vale destacar que cada uma destas relações é fruto de um relacionamento temporal entre dois ou mais objetos mulsemídia. Além disso, a região de reprodução do conteúdo audiovisual também exibe ícones que representam efeitos

sensoriais sincronizados temporalmente com as mídias tradicionais. Por exemplo, um efeito de calor é ilustrado por um ícone de fogo, atrelado ao início do primeiro vídeo.

## 5.2 Exportação para NCL 4.0

Dado que o modelo MultiSEM, utilizado pelo STEVE 2.0 para especificar as aplicações mulsemídia apresenta uma modelagem simples, principalmente, na definição das relações temporais, a exportação das aplicações criadas na ferramenta para outro modelo é facilitada. Dessa forma, um projeto criado no STEVE 2.0 pode facilmente ser exportado para aplicações NCL 4.0.

A partir do modelo MultiSEM, o sistema pode gerar um documento mulsemídia em NCL 4.0, transformando as entidades do modelo em elementos da linguagem NCL 4.0. Para isto, a API ANA [16] foi estendida para exportar projetos STEVE 2.0 em documentos NCL 4.0, representando elementos NCL como classes Java. A API ANA é um metamodelo criado especificamente para representar documentos NCL em ambientes orientados a modelos, aprimorando a manipulação do código NCL. O editor também usa métodos ANA para escrever os documentos NCL. A extensão na API ANA foi necessária para suportar o nó de efeito sensorial presente em NCL 4.0, pois a API dava suporte apenas a elementos da linguagem NCL 3.0.

## 6 Limitações do STEVE

O STEVE enfrenta atualmente algumas limitações, como a ausência de interação multimodal e a falta de suporte para múltiplos usuários. A falta desse primeiro quesito significa que os criadores de conteúdo estão restritos a interações unidimensionais, geralmente limitadas ao uso de teclas em um controle remoto. Isso pode reduzir a capacidade de transmitir narrativas complexas e também limitar a eficácia da comunicação e o impacto do conteúdo produzido.

Já as implicações causadas pela inexistência de suporte multiusuário impactam em questões de personalização, onde não é possível criar aplicações que contemplem os gostos pessoais dos telespectadores, ao passo que também dificulta a criação de níveis de acesso, onde cada usuário ficaria restrito ao seu perfil. Do ponto de vista de autoria, essa carência impede a colaboração coletiva, a combinação de habilidades e a troca dinâmica de ideias por parte dos desenvolvedores, o que também dificulta a revisão e o aprimoramento contínuo do conteúdo, uma vez que eles exigem adaptações rápidas e iterações constantes.

Em resumo, as limitações do STEVE em relação à interação multimodal e ao suporte multiusuário não apresentam consequências substanciais na criação de conteúdo multimídia, mas limitam a capacidade de transmitir mensagens complexas e afetam a colaboração entre criadores e a elaboração de conteúdo personalizado. Para atender às demandas da TV 3.0 que será mais interativa e colaborativa que a versão atual,

um dos trabalhos futuros em relação a ferramenta é movê-la em direção à superação dessas limitações.

## 7 Conclusão

O presente artigo propôs o emprego do STEVE 2.0, uma ferramenta de autoria mulsemídia com modelo de sincronização baseado em eventos, que oferece uma interface gráfica a partir de linhas do tempo, para facilitar a autoria de aplicações multissensoriais para a TV 3.0. Considerando sua facilidade e intuitividade para uso, STEVE tem potencial incentivar a adoção dessa tecnologia em larga escala. A ferramenta permite a exportação automática das aplicações multissensoriais criadas em documentos NCL 4.0, tecnologia que foi selecionada para a camada de codificação de aplicações do novo sistema de TV.

Já para a implementação de sistemas multissensoriais como um todo, os desafios incluem questões tecnológicas, como a sincronização em tempo real das mídias com os efeitos, além dos desafios relacionados à integração de efeitos subjetivos e de difícil aplicação, como aqueles relacionados ao paladar e ao olfato. As questões econômicas perpassam transformações na indústria, seja na criação de aparelhos e hardwares que suportem este tipo de comunicação, seja na parte de produção de conteúdo que deve encontrar equilíbrio entre automação e qualidade. Por fim, desafios sociais incluem a distribuição e acesso desigual das novas tecnologias.

A TV 3.0 é a futura geração do sistema de televisão digital brasileiro que conta com integração mulsemídia. Nesse contexto, é fundamental encontrar um equilíbrio entre inovação e difusão dessas novas técnicas. A evolução da televisão não se trata apenas de tecnologia, mas também de criar conexões emocionais mais profundas com o público e de enriquecer a forma como as histórias são contadas e compartilhadas. Desse modo, fornecer ferramentas inclusivas, acessíveis e de fácil emprego na prática, torna-se essencial para experimentar este novo mundo audiovisual que se vislumbra.

## Agradecimentos

Os autores agradecem o apoio do Ministério da Comunicações e das agências financiadoras CAPES (88887.668167/2022-00), RNP, CNPq e FAPERJ.

## Referências

- [1] Raphael Abreu, Douglas Mattos, Joel AF dos Santos, and Débora C Muchalut-Saade. 2019. Semi-automatic synchronization of sensory effects in mulsemimedia authoring tools. In *Proceedings of the 25th Brazilian Symposium on Multimedia and the Web*. 201–208.
- [2] James F Allen. 1983. Maintaining knowledge about temporal intervals. *Commun. ACM* 26, 11 (1983), 832–843.
- [3] Willians Cerozzi Balan and Doutor do Curso de Radialismo-Unesp. 2012. Um breve olhar pela evolução da TV no Brasil. *Revista Produção Profissional, São Paulo* 1, 124 (2012), 12.
- [4] André Barbosa Filho. 2007. A televisão digital no Brasil: do sonho à realidade. *Comunicação & Sociedade* 29, 48 (2007), 27–47.

- [5] Samia Bouyakoub and Abdelkader Belkhir. 2011. SMIL builder: An incremental authoring tool for SMIL documents. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 7, 1 (2011), 1–30.
- [6] M. Cecelia Buchanan and Polle T. Zellweger. 1993. Specifying Temporal Behavior in Hypermedia Documents. In *Proceedings of the ACM Conference on Hypertext* (Milan, Italy) (ECHT '92). Association for Computing Machinery, New York, NY, USA, 262–271. <https://doi.org/10.1145/168466.171513>
- [7] DCA Bulterman, Lynda Hardman, J Jansen, KS Mullender, and L Rutledge. 1998. A Graphical Interface for creating and playing SMIL documents. In *Proceedings of the EMMSEC 98 conference*. Elsevier, Amsterdam, NL, 11.
- [8] Bumsuk Choi, Eun-Seo Lee, and Kyoungro Yoon. 2011. Streaming media with sensory effect. In *2011 International Conference on Information Science and Applications*. IEEE, IEEE Xplore, Jeju, South Korea, 1–6. <https://doi.org/10.1109/ICISA.2011.5772390>
- [9] Casa Civil, LUIZ INÁCIO LULA DA SILVA, APARECIDA GONÇALVES, et al. 2023. Decreto nº 11.485, de 6 de abril de 2023. *Diário Oficial da União (DOU)* 1, 1 (2023), 2.
- [10] Fabien Danieau, Jérémie Bernon, Julien Fleureau, Philippe Guillotel, Nicolas Mollet, Marc Christie, and Anatole Lécuyer. 2013. H-Studio: an authoring tool for adding haptic and motion effects to audiovisual content. In *Proceedings of the adjunct publication of the 26th annual ACM symposium on User interface software and technology*. ACM Digital Library, New York, NY, USA, 83–84. <https://doi.org/10.1145/2508468.2514721>
- [11] Raphael Silva de Abreu, Douglas Mattos, Joel dos Santos, Gheorghita Ghinea, and Débora Christina Muchaluat-Saade. 2020. Toward content-driven intelligent authoring of mulsemmedia applications. *IEEE Multi-Media* 28, 1 (2020), 7–16.
- [12] Douglas Paulo de Mattos. 2021. *An Approach for Authoring Mulsemmedia Applications Based on Events*. Ph.D. Dissertation. Universidade Federal Fluminense - UFF.
- [13] Douglas Paulo de Mattos and Débora C Muchaluat-Saade. 2018. Steve: A hypermedia authoring tool based on the simple interactive multimedia model. In *Proceedings of the ACM Symposium on Document Engineering 2018*. ACM Digital Library, Halifax, Canada, 1–10.
- [14] Douglas P de Mattos, Débora C Muchaluat-Saade, and Gheorghita Ghinea. 2020. An approach for authoring mulsemmedia documents based on events. In *2020 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, IEEE Xplore, Big Island, USA, 273–277.
- [15] Maria Cristina Ferreira De Oliveira, Marcelo Augusto Santos Turine, and Paulo Cesar Masiero. 2001. A statechart-based model for hypermedia applications. *ACM Transactions on Information Systems (TOIS)* 19, 1 (2001), 28–52.
- [16] Joel AF dos Santos, Julia V da Silva, Renan R Vasconcelos, Wagner Schau, Cláudia Werner, and Débora CM Saade. 2012. aNa: API for NCL authoring. In *Anais Estendidos do XVIII Simpósio Brasileiro de Sistemas Multimídia e Web*. SBC, 63–66.
- [17] Joel André Ferreira dos Santos. 2016. *Multimedia Document Validation Along Its Life Cycle*. Ph.D. Dissertation. Universidade Federal Fluminense (UFF), Niterói. <https://www.ic.uff.br/teses-e-dissertacoes/>
- [18] Natalia C Fernandes, Vitor LG Mota, Amanda Beatriz C dos Santos, Gabriel S Vencioneck, Dianne SV Medeiros, Diogo MF Mattos, Leni Joaquim de Matos, et al. 2022. Recepção Móvel na TV 3.0: um caso de estudo com a tecnologia Advanced ISDB-T. *SET EXPO Proceedings* 8, 1 (2022), 3.
- [19] João Freire Filho. 2007. História da Televisão: Teoria e Prática. *IV Encontro dos Núcleos de Pesquisa do Intercom*. Disponível em: <http://reposcom.portcom.intercom.org.br/>. Acedido a 1 (2007), 19.
- [20] Gheorghita Ghinea, Christian Timmerer, Weisi Lin, and Stephen R Gulliver. 2014. Mulsemmedia: State of the art, perspectives, and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 11, 1s (2014), 1–23.
- [21] Victor Giroto, Erin Walker, and Winslow Burleson. 2019. CrowdMUSE: Supporting crowd idea generation through user modeling and adaptation. In *Proceedings of the 2019 on creativity and cognition*. ACM Digital Library, New York, NY, USA, 95–106.
- [22] Céline Jost, Brigitte Le Pévédic, Justin Debloos, and Gérard Uzan. 2022. Interactions in Multisensory Experiences: Toward a New Timeline Metaphor. In *Proceedings of the 2nd Workshop on Multisensory Experiences-SensoryX'22*. SBC.
- [23] M. Josué, R. Abreu, F. Barreto, D. Mattos, G. Amorim, J. dos Santos, and D. Muchaluat-Saade. 2018. Modeling sensory effects as first-class entities in multimedia applications. In *Proceedings of the 9th ACM Multimedia Systems Conference*. ACM, ACM Digital Library, New York, NY, USA, 225–236. <https://doi.org/10.1145/3204949.3204967>
- [24] Marina Ivanov Pereira Josué. 2021. *Preparação de Objetos de Mídia e Efeitos Sensoriais para Formatação de Documentos Mulsemídia*. Ph.D. Thesis. Universidade Federal Fluminense. <https://www.ic.uff.br/teses-e-dissertacoes/>
- [25] S. Kim. 2013. Authoring multisensory content. *Signal Processing: Image Communication* 28, 2 (2013), 6. <https://doi.org/10.1016/j.image.2012.10.011>
- [26] S. Kim and J. Han. 2004. *Text of white paper on MPEG-V*. International Organization for Standardization. <https://www.mpeg.org/whitepapers/>
- [27] Ina Koch, Wolfgang Reisig, and Falk Schreiber. 2010. *Modeling in systems biology: the Petri net approach*. Vol. 16. Springer science & business media, New York, NY, USA.
- [28] William Mansque. 2022. *Como a TV evoluiu no Brasil nos últimos 50 anos e o que vem pela frente*. Gaúcha Zero Hora. <https://gauchazh.clicrbs.com.br/cultura-e-lazer/tv/noticia/2022/02/como-a-tv-evoluiu-no-brasil-nos-ultimos-50-anos-e-o-que-vem-pela-frente-ckzr67evs00f015pkeeobrn8.html>
- [29] Douglas P Mattos and Débora C Muchaluat-Saade. 2018. MultiSEM: A mulsemmedia model for supporting the development of authoring tools. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*. ACM Digital Library, Salvador, BA, Brasil, 109–116.
- [30] Fórum SBTVD-T. 2023. *Background – Terrestrial TV Evolution in Brazil*. Fórum SBTVD-T. [https://forumsbtvd.org.br/tv3\\_0/#panel-62d193f2420ab](https://forumsbtvd.org.br/tv3_0/#panel-62d193f2420ab)
- [31] Sang-Ho Shin, Keum-Sook Ha, Han-O Yun, and Yoon-Seok Nam. 2016. Realistic media authoring tool based on MPEG-V international standard. In *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, IEEE Xplore, Vienna, Austria, 730–732. <https://doi.org/10.1109/ICUFN.2016.7537133>
- [32] Luiz Fernando Gomes Soares. 2009. *Programando Em NCL*. Campus, Rio de Janeiro, Brazil. 360 pages.
- [33] Guido Van Rossum, Jack Jansen, K Sjoerd Mullender, and Dick CA Bulterman. 1993. CMIFed: a presentation environment for portable hypermedia documents. In *Proceedings of the first ACM international conference on Multimedia*. ACM Digital Library, New York, NY, USA, 183–188.
- [34] M. Waltl, B. Rainer, C. Timmerer, and H. Hellwagner. 2013. An end-to-end tool chain for Sensory Experience based on MPEG-V. *Signal Processing: Image Communication* 28, 2 (2013), 136–150. <https://doi.org/10.1016/j.image.2012.10.009>