# Acessibilidade na TV 3.0 Brasileira a partir de mídias de legenda, glosa e áudio descrição.

Richelieu R. A. Costa *
richelieu.costa@lavid.ufpb.br

Derzu Omaia *
derzu@lavid.ufpb.br

Tiago M. U. Araújo *
tiagomaritan@lavid.ufpb.br

Jóison O. Pereira *
joison.pereira@lavid.ufpb.br

Anderson S. Coutinho *
anderson.coutinho@lavid.ufpb.br

Miguel P. S. Cruz
miguel.cruz@lavid.ufpb.br*

Victoria M. Pontes *
victoria.pontes@lavid.ufpb.br

Matheus M. Barbosa *
matheus.mendonca@lavid.ufpb.br

Abner S. Silva *
abner.souza@lavid.ufpb.br

Guido L. S. Filho *
guido@lavid.ufpb.br

* Univ. Federal da Paraíba
João Pessoa, PB, Brasil

## Abstract

The TV 3.0 project is under development, managed by the Forum of the Brazilian Digital Terrestrial TV System (SBTVD), with the participation of researchers from academia and industry. There was a Call for Proposals for this development, resulting in 36 responses from 21 organizations worldwide. The proposals were evaluated and examined, and the SBTVD Forum forwarded its recommendations on selecting candidate technologies to the Brazilian Ministry of Communications. Currently, researchers are working on elaborating specifications for video encoding, audio encoding, and subtitles. This work analyzed the accessibility requirements for subtitles, glosses, and audio description media. Solutions were presented for their implementation by proposing an extension to the CCWS server API and a system architecture, which was implemented as a proof of concept.

***Keywords:*** Application coding, TV 3.0, Accessibility, SBTVD, gloss, sign language, audio description, closed caption

## 1 Introdução

A televisão (TV) tem um papel importante na sociedade brasileira. Além de fornecer entretenimento individual e coletivo, ela também atua como elemento informativo, educacional e de interação social [15], [16]. Dessa forma, qualquer alteração a ser proposta ao modo de assistir TV do brasileiro tem que ser cuidadosamente analisada. O Projeto da TV 3.0 está em desenvolvimento e é gerenciado pelo Fórum do SBTVD, criado para assessorar o Governo Brasileiro em políticas e questões técnicas relacionadas à aprovação de inovações técnicas, especificações, desenvolvimento e implementação do SBTVD. Sendo assim, as pesquisas realizadas abrangem não somente as tecnologias envolvidas, mas também o impacto social.

A TV chegou no Brasil na década de 1950, com imagens em preto e branco, e em 1972 as primeiras transmissões em cores foram realizadas [11]. Em 2007, um padrão de transmissão terrestre, com tecnologia nacional, foi especificado em conjunto com o middleware Ginga, o qual permitia que o usuário interagisse com a TV através de aplicativos transmitidos pela emissora. Em 2021, uma nova especificação chega ao mercado, adicionando novos recursos, entre eles, aumento na memória de armazenamento persistente, novos decodificadores de áudio e vídeo, e um servidor *webservices* (*Common Core Web Service* - CCWS). Esta versão da TV ficou conhecida como profile D, ou DTVPlay. O recurso do CCWS permite que dispositivos externos comuniquem-se e interajam com a TV através de uma API web RESTful. Em 2020 foi iniciada uma fase de desenvolvimento de um novo padrão, chamado de TV 3.0, através de uma uma chamada por proposta das tecnologias que poderiam ser adotadas na TV 3.0. Em 2021 os proponentes forneceram documentos, softwares e equipamentos para que suas tecnologias propostas pudessem ser avaliadas por laboratórios de testes acadêmicos.

Atualmente a TV 3.0 está na sua terceira fase de desenvolvimento. Nesta fase, testes e avaliações sobre a camada física e de vídeo estão sendo realizados e está sendo desenvolvido um multiplexador e demultiplexador de referência. Na camada de codificação de aplicações, a maioria dos requisitos inovadores propostos estão sendo analisados e parcialmente implementados por grupos de pesquisa em universidades, visto que, estes requisitos não foram completamente resolvidos nas etapas anteriores. Os requisitos avançados incluem

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.

ISSN 2596-1683

o suporte a TV baseada em aplicativos, conteúdo audiovisual imersivo, interação multimodal, efeitos sensoriais, perfis multi-usuários, aferição de audiência, convergência IP, acessibilidade, extensibilidade, entre outros.

Este trabalho endereça os requisitos de acessibilidade que estão sendo estudados e desenvolvidos no âmbito no projeto TV 3.0, mais especificamente, o suporte a legendas ocultas (do inglês, *closed caption*), o suporte a língua de sinais, e o suporte a áudio descrição. O Censo do IBGE de 2010, aponta que o Brasil tem 2,16 milhões de brasileiros surdos, e 6,6 milhões cegos [8]. Dessa forma, o desenvolvimento dessas tecnologias vai permitir que esse público posso interagir com a TV de forma mais inclusiva.

Esse requisitos são essenciais, porque geralmente as Tecnologias da Informação e Comunicação (TIC), quando são projetadas, dificilmente levam em conta os requisitos e necessidades das pessoas com deficiência [7]. O suporte para línguas de sinais, por exemplo, é raramente explorado nessas tecnologias. Na TV, por exemplo, o suporte a línguas de sinais é, em geral, limitado a uma janela com um intérprete de língua de sinais, apresentada juntamente com o vídeo original do programa (*wipe*). Essa solução possui altos custos operacionais para geração e produção (câmeras, estúdio, equipe, etc.) dos conteúdos, necessita de intérpretes humanos em tempo integral, o que acaba restringindo seu uso a uma pequena parcela da programação. Essas dificuldades resultam em uma grande barreira para a comunicação com outras pessoas, o acesso a informações, a aquisição de conhecimentos, dentre outros.

Para endereçar essa questão, neste trabalho, está sendo proposta a disponibilização da janela de língua de sinais utilizando avatares 3D e tradução automática para línguas de sinais, com o intuito de reduzir os custos operacionais envolvidos e viabilizar uma oferta maior de conteúdos com língua de sinais na programação da TV 3.0 no Brasil. Nesta solução, o radiodifusor transmitirá um fluxo contendo uma sequência de glosas (representação textual na gramática de língua de sinais), que será convertido para Língua Brasileira de Sinais (Libras) na própria TV ou num dispositivo de segunda tela, utilizando o componente de síntese de sinais (*player*) e os avatares 3D do VLibras[1].

O restante do artigo está organizado da seguinte forma. Na seção 2, é apresentado o referencial teórico, que discutirá os principais conceitos relacionados a este trabalho, tais como legenda oculta, Libras e o VLibras. Nas Seções 3 e 4, serão apresentadas a metodologia e as provas de conceito das soluções desenvolvidas. Na Seção 5, serão apresentados os principais desafios relacionados a este tipo de solução na TV 3.0. Por fim, na Seção 6 serão apresentadas as conclusões e propostas de trabalhos futuros.

[1]O VLibras é um conjunto de ferramentas gratuitas e de código aberto que traduz conteúdos digitais (texto, áudio e vídeo) em Português para Língua Brasileira de Sinais (Libras), para permitir que computadores, dispositivos móveis e plataformas Web sejam mais acessíveis para as pessoas surdas [4].

## 2 Referencial Teórico

Nesta seção são apresentados aspectos técnicos sobre legendas, Língua de Sinais e o VLibras.

### 2.1 Legendas

As legendas ocultas (*closed caption*), ou apenas legendas, são mídias em formato texto, que, geralmente, transcrevem o que está sendo falado no vídeo para que possa ser lida pelos telespectadores. É bastante utilizada em situações nas quais o áudio do vídeo está em um idioma em que o telespectador não é fluente, desta forma, uma ou mais legendas são transmitidas, possibilitando a leitura da legenda em outros idiomas.

Na fase 2 do projeto da TV 3.0, foi decidido que seria utilizado o padrão IMSC1[13], do *Advanced Television System Committee* (ATSC) A/343A [3] para a codificação e transmissão de legendas. A transmissão de glosa e mensagens de alertas de emergências também são codificadas nesse formato de legenda. Este padrão será brevemente apresentado nesta seção.

O documento normativo da ATSC A/343A [3] define a tecnologia necessária para legendas em transportes ROUTE-DASH e MMT. Isso inclui o conteúdo, o empacotamento e a sincronização. A tecnologia é baseada em SMPTE Timed Text (SMPTE-TT), conforme definido em SMPTE 2052-1 [1].

O SMPTE-TT é [1] compatível com tabelas de símbolos e idiomas em todo o mundo, suporta entrega de glifos de imagem, está em uso hoje por vários "silos de entrega de mídia", incluindo serviços de radiodifusão entregues pela Internet. Este padrão é complexo e vai além do necessário para atender os requisitos de legendas. Diante disso, um subconjunto mais simples é desejável para implementação prática. Portanto, o "*TTML Profiles for Internet Media Subtitles and Captions* (IMSC1)" [13] do W3C foi selecionado para necessidades como broadcast e entrega de broadband. O padrão TTML IMSC1 do W3C é utilizado para fornecer legendas como um componente separado do vídeo, que pode ser transmitido por broadcast, conforme descrito no documento normativo ATSC A/343 [3].

Neste padrão é utilizado um subset do padrão TTML, o qual consiste em um arquivo XML com diversas configurações possíveis para a legenda, tais como posicão, cor, fonte, tempo de apresentação, sincronismo, emojis e imagens.

Quando há legendas disponível, o conteúdo é especificado e transmitido usando um ou mais arquivos ISO BMFF, cada um contendo um ou mais documentos XML. Os documentos XML estão em conformidade com o IMSC1 conforme restrito e estendido no documento normativo ATSC A/343 [3]. Cada arquivo contém apenas um conjunto de "texto cronometrado" correspondente a um conjunto de "sinalização" de metadados.

Para transmitir um conteúdo previamente existente, os segmentos de captions em ISO BMFF (ou seja, documentos

IMSC1) precisam ter uma duração relativamente curta. Isso é necessário para permitir que os decodificadores participem de uma transmissão em andamento, adquiram e apresentem o conteúdo da captions simultaneamente com o conteúdo do programa de áudio e vídeo. O tempo de aquisição e apresentação das legendas (se presentes nesse momento) deverá ser da ordem do tempo de aquisição e apresentação de vídeo e áudio. A duração do documento IMSC1, portanto, normalmente varia de 1/3 a 3 segundos. Documentos TTML mais longos, embora sejam mais eficientes, podem resultar em atrasos questionáveis na primeira apresentação do conteúdo da legenda [13].

### 2.2 Língua de Sinais

A língua de sinais é uma forma de comunicação gestual que possibilita a comunicação com pessoas surdas. Libras é a linguagem brasileira de sinais, utilizada para exercer a comunicação para deficientes auditivos, tendo um papel importante para a integração dessas pessoas, tornando sua aplicação essencial nas plataformas de comunicação [6].

Glosa é um texto, descritivo em libras, que obedecem as regras gramaticais dessa linguagem, tem por objetivo facilitar a compreensão, também tradução por interpretes, além da comunicação textual entre os fluentes de sinais [6].

### 2.3 VLibras

A Suíte VLibras é um conjunto de ferramentas de código aberto que traduzem automaticamente conteúdos digitais de língua portuguesa para Libras, tornando informações mais acessíveis para pessoas surdas em computadores, dispositivos móveis, TVs, plataformas Web, entre outros [4].

O projeto é resultado de uma parceria entre o Ministério da Gestão e Inovação em Serviços Públicos (MGISP), por meio da Secretaria de Governo Digital (SGD), o Ministério dos Direitos Humanos e da Cidadania (MDHC), por meio da Secretaria Nacional dos Direitos da Pessoa com Deficiência (SNDPD), e a Universidade Federal da Paraíba (UFPB), através do Laboratório de Aplicações de Vídeo Digital (LAVID) [4].

Os principais componentes do VLibras são:

- VLibras-Plugin e VLibras-Widget: extensões de navegador web que permitem que textos selecionados em páginas Web sejam traduzidos automaticamente para Libras e reproduzidas através de um avatar 3D;
- VLibras-Móvel: aplicação cliente do VLibras para dispositivos móveis (compatível com os sistemas Android e iOS);
- VLibras-Desktop: ferramenta utilizada para traduzir automaticamente textos selecionados em programas executados em computadores pessoais para Libras;
- VLibras-Vídeo: um portal que permite tradução para Libras de trilhas de áudio e legendas associadas ao vídeo. (Atualmente só disponível para vídeos legendados e sob demanda, pois o tempo para renderizar o vídeo com a tradução é maior que o tempo do vídeo.) ;

Outra parte do VLibras são os serviços de retaguarda (ou *backend*), denominados VLibras-services, que realizam a tradução automática para os outros componentes (ou ferramentas) e armazenam animações 3D dos sinais em Libras que são utilizadas para renderizar os conteúdos acessíveis após a tradução. Atualmente, o dicionário de sinais tem aproximadamente 21 mil sinais em Libras, uma maiores bases de dados deste tipo no mundo [4].

Por fim, tem-se uma ferramenta colaborativa, denominada WikiLibras, que permite que voluntários participem do processo de construção e expansão do dicionário de sinais [12].

## 3 Metodologia

Este artigo foca na metodologia, progresso, propostas e recentes conquistas do grupo de pesquisa da Universidade Federal da Paraíba (UFPB) em relação aos requisitos do desenvolvimento de aplicações na área de acessibilidade da TV 3.0. A equipe é composta por 10 pesquisadores, entre doutores, mestres e graduandos, o trabalho, nesta fase 3, começou em Abril de 2023.

Como forma de colaborar ativamente com a metodologia de investigação, os Módulos Técnicos e de Mercado do Fórum SBTVD decidiram em conjunto uma priorização de requisitos para determinar a sequência de estudos para a equipe de pesquisa e desenvolvimento (P&D). Além disso, o Grupo de Trabalho (GT) de Codificação de Aplicações do Fórum especificou as diretrizes iniciais sobre como abordar cada requisito, com base nos resultados da avaliação da Fase 2 e na experiência do GT na padronização do middleware DTV. Finalmente, o GT especificou um total de 7 casos de uso a serem prototipados, com o objetivo de validar as soluções de P&D e demonstrar publicamente as novas funcionalidades da TV 3.0.

A equipe de P&D incorporou todas as contribuições do Fórum SBTVD na sua metodologia, permitindo um progresso consistente em determinados requisitos e apresentando já os resultados iniciais dos estudos.

O trabalho desenvolvido na UFPB, atualmente, foca no caso de uso 2 (AP-uc-2) especificado pelo GT. O qual trata sobre o Fornecimento de conteúdos sincronizados em vários dispositivos, incluindo acessibilidade.

Desta forma, estão sendo investigadas e propostas soluções para a transmissão e recebimento de legendas, glosas, e áudio descrição em dispositivos externos a TV de forma sincronizada. Para isso, é proposto um sistema onde a TV gerencia esse processo de comunicação com os dispositivos externos através do servidor CCWS.

### 3.1 Visão geral da solução

A arquitetura proposta do sistema é centrada na distribuição e sincronismo das mídias de acessibilidade a partir do servidor CCWS. A televisão recebe o sinal digital da TV 3.0, demultiplexa e decodifica este. Desta forma, possibilita que mídias especificas sejam acessadas e distribuídas pelo CCWS. Uma API Rest, que ainda está em fase de proposição, foi estendida a partir da existente na TV2.5. Ela fornece rotas que permitem a um cliente, externo a TV e conectado na mesma rede local, acessar essas mídias, desde que esteja autenticado e conectado na TV.

Deste modo, as aplicações clientes podem requisitar e acessar as mídias de glosa, legendas, e áudio descrição. A partir disso, diversos cenários podem ser explorados. No caso de telespectadores deficientes auditivos, a glosa recebida, pode ser exibida no formato de Língua de Sinais no dispositivo do usuário, sem sobrepor ao vídeo que está sendo exibindo na tela da TV. Para o caso de exibição de legendas, diferentes telespectadores podem acessar legendas no idioma que preferirem, possibilitando que pessoas diferentes recebam conteúdos diferentes. E o cliente de áudio descrição pode receber o áudio em seu celular e ouvi-lo através de fones de ouvido, sem que os demais telespectadores precisem ouvir a áudio descrição.

Isso permite uma personalização do conteúdo, de forma simultânea e não impositiva, visto que, cada usuário pode ter sua própria personalização no seu dispositivo pessoal, sem interferência dos demais.

A Figura 1 apresenta essa arquitetura e demonstra os três cenários apresentados. É possível ver o servidor CCWS na TV entregando os 3 conteúdos de mídias de acessibilidade aos dispositivos móveis, através de uma rede wi-fi. Cada dispositivo recebe a sua mídia a reproduz de acordo com o seu tipo. Um usuário da acessibilidade consegue visualizar ou ouvir o conteúdo recebido nos dispositivos.

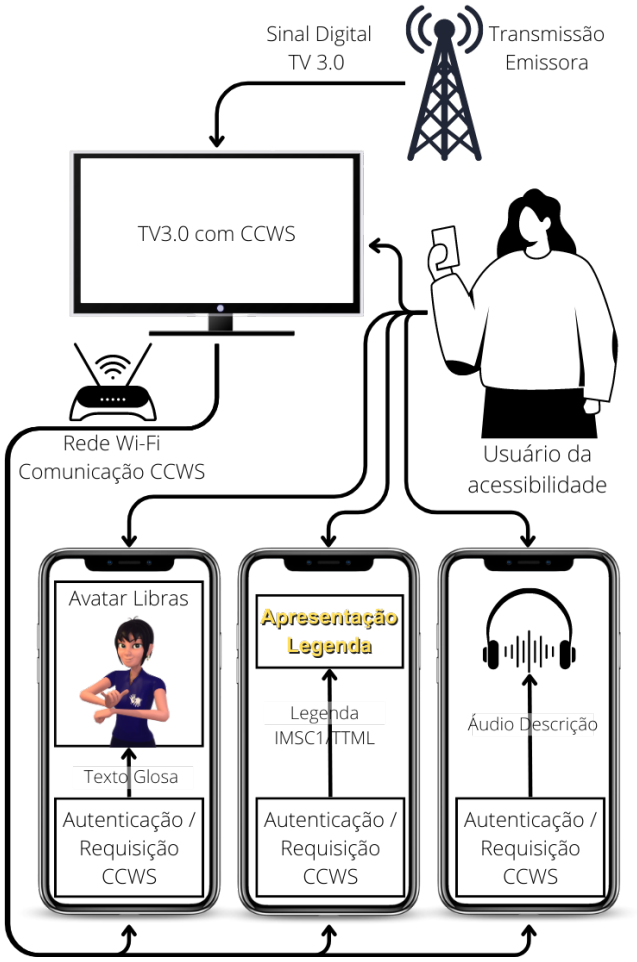


**Figure 1.** Arquitetura Geral Acessibilidade na TV 3.0

### 3.2 API proposta CCWS

Foi proposto um conjunto de novas APIs para acessibilidade para o CCWS, estendendo as rotas já existentes no CCWS da TV 2.5. Essas novas APIs possibilitam o encaminhamento, em tempo real, de legendas e glosa para dispositivos móveis através de sockets, websockets ou HTTP, em rede local e de forma síncrona. Para validação da proposta foi realizada uma implementação parcial de testes do Ginga CCWS, para um ambiente simulado.

As novas rotas propostas estendem a rota "8.5.1 Acesso direto a um stream" da ABNT NBR 15606-11 [5], possibilitando a requisição específica para diversas mídias, entre elas as mídias de acessibilidade e legendas. A rota proposta é apresentada a seguir:

**Listing 1.** Rota acesso mídias

```
http(s)://<host>/dtv/current-service/stream/
<alias>[?id=<id>][&protocol=<protocol>]
```

O *alias* da rota, indica o tipo de mídia que será requisitada, podendo ter os valores apresentados na Tabela 1. As mídias de caption e glosa estão no formato XML TTML, que é o formato utilizado pelo IMSC1 [13] para a codificação de legendas. Para a transmissão do fluxo de áudio descrição foi proposto usar o formato DASH [9] e não mais o RTSP, como era proposto no CCWS da TV2.5 [5]. Essa mudança ocorreu porque o sinal recebido pela TV 3.0 será no protocolo ROUTE/DASH, desta forma, seria mantido o formato original dos dados. Para a transmissão dos arquivos TTML, foi proposto o envio via sockets e/ou websockets, devido a natureza textual dos dados, o que torna mais simples sua recepção pela aplicação cliente. Caso fosse transmitido em DASH a aplicação cliente teria que decodificar o DASH para extrair o TTML. Além disso, este formato já era utilizado no CCWS da TV2.5, para a recepção de stream events DSMCC [5]. Já para a transmissão de glosa decidiu-se também utilizar

o padrão TTML, contudo, isto não é especificado no padrão IMSC1.

**Table 1.** Nomes das mídias, seus apelidos e tipos

| Nome da Mídia | Alias | Mime Type |
|---|---|---|
| Áudio descrição | audiodescription | application/dash+xml |
| Glosa | gloss | application/ttml+xml |
| Legendas | captions | application/ttml+xml |

A transmissão dessas mídias para o cliente é realizada através de sockets ou websockets para as mídias de legenda e glosa, e o protocolo do *socket* pode ser especificado através do parâmetro de *query <protocol>*. Já a áudio descrição é transmitida via protocolo HTTP.

O parâmetro de *query <id>* possibilita a especificação do código de identificação de uma mídia específica, caso haja mais de um mídia do mesmo tipo. Se omitido, a mídia padrão, para este tipo, será utilizada.

O retorno da rota é em formato JSON e fornece as informações necessárias para que o cliente se conecte ao servidor. A Listagem 2 apresenta o formato do retorno.

**Listing 2.** JSON de retorno da rota de acessibilidade

```
{
    "handle": "<handle>",
    "url": "<streamUrl>"
}
```

O *handle* é um identificador da requisição que deve ser utilizado para liberar o recurso de transmissão da mídia após sua utilização.

A URL contém a informação para abertura do socket, websocket, ou o *link* para o arquivo DASH [9].

A rota proposta também poderá ser utilizada para recepção das mídias principais de áudio e vídeo em formato DASH, contudo, não estamos apresentando seus *alias* neste trabalho, devido ao escopo ser específico para a área de acessibilidade.

## 4 Prova de conceito

Para validação e demonstração da proposta foi realizada uma implementação parcial do Ginga CCWS e uma aplicação de teste para um ambiente simulado.

No servidor CCWS foi implementada a rota de acesso às mídias de acessibilidade, seguindo a especificação proposta na Listagem 1. Contudo, as mídias disponibilizadas são simuladas, visto que, não está sendo implementado um sistema completo de transmissão e recepção de TV terrestre 3.0.

A aplicação de testes foi desenvolvida em HTML+ Javascript e possui uma versão *desktop*, que pode ser executada na TV e uma versão *mobile* para ser executada em dispositivos móveis. Independente da forma de apresentação a aplicação possui os recursos de exibição de Língua de Sinais, exibição de legenda e de reprodução de áudio descrição.

Na versão *desktop*, ainda é exibida a mídia de vídeo, a qual não fazia parte do escopo inicial da pesquisa, visto que é a mídia principal da TV. Contudo, para uma melhor apresentação do sistema proposto foi necessário simular sua recepção e apresentação. Ela foi reproduzida a partir de um arquivo local da aplicação, sem transmissão pelo CCWS.

Cada módulo requisita ao servidor a sua mídia específica e inicia sua recepção e tratamento para reproduzir a mídia de forma apropriada.

A seguir, o CCWS e os módulos da aplicação de Língua de Sinais, de legendas e de áudio descrição são detalhados.

### 4.1 CCWS - Rota de acessibilidade

A implementação parcial do CCWS, para o presente trabalho, tem como principal objetivo permitir que aplicações não locais, executadas em dispositivos externos, possam acessar mídias de acessibilidade transmitidas por uma emissora de TV 3.0.

A rota proposta na Listagem 1 foi implementada em um servidor Node.js. As mídias a serem disponibilizadas foram preparadas de forma que as legendas e glosas ficassem com seu conteúdo equivalente. Para isso, inicialmente, foi gerado manualmente, a legenda para um determinado vídeo, seguida de sua tradução para glosa utilizando o tradutor do VLibras [14]. Para simular um ambiente de transmissão real, no qual as legendas e glosas são transmitidas continuamente, estas foram segmentadas em partes de 2 segundos, na qual, cada uma foi armazenada em um arquivo TTML diferente. Dessa forma, o conteúdo é transmitido de maneira simultânea e síncrona, a cada 2 segundos, para os clientes conectados através de sockets ou websockets. Ao final da transmissão dos arquivos, o servidor CCWS reinicia o ciclo, voltando a transmitir os primeiros arquivos novamente.

O fluxo de áudio descrição é disponibilizado através de uma URL HTTP e ainda não foi sincronizado com a legenda e glosa na fase atual do projeto.

### 4.2 Módulo Legendas

A representação visual das legendas em formato IMSC é realizada por meio da linguagem de programação JavaScript (JS) e da biblioteca de código aberto imscJS[10]. Essa biblioteca desempenha um papel fundamental ao interpretar o conteúdo presente nos arquivos de legendas. Esses arquivos seguem o padrão TTML, alinhado com as especificações do perfil IMSC definido pelo W3C. Sendo assim, a biblioteca permite a exibição das legendas na aplicação com a formatação apropriada. Por meio dessa abordagem, abre-se a possibilidade de personalizar diversos aspectos visuais na representação gráfica delas, como cores, tamanhos e posicionamento. Adicionalmente, é viável incorporar elementos complementares em sua exibição, como emojis e imagens.

Para a versão *desktop* da aplicação, o sincronismo com o vídeo é realizado deslocando, no tempo, o vídeo para o momento de apresentação da primeira legenda recebida. Esse

momento é obtido das marcações temporais presentes no arquivo TTML.

Já o sincronismo entre os diversos clientes é realizado através servidor CCWS, o qual envia simultaneamente para todos os clientes, o mesmo conteúdo. As marcações temporais do arquivo TTML são utilizadas pela biblioteca imscJS[10] para exibir a legenda no momento e pelo tempo corretos, proporcionando uma experiência de visualização coerente e sincronizada.

### 4.3 Módulo Língua de sinais

O módulo de língua de sinais recebe do servidor CCWS as glosas através de um websocket. Para a representação em formato de língua de sinais foi realizada a integração da aplicação com o VLibras Widget [4], uma ferramenta que possui um avatar 3D que reproduz as glosas em formato de língua de sinais.

O Widget [4] foi ajustado para que funcionasse de forma *offline*, sem necessidade de internet. Para isso, o dicionário de animações e os arquivos JavaScript do Widget foram hospedados em um servidor HTTP, junto ao servidor CCWS. Desta forma, dispositivos conectados a TV através de uma rede local, mesmo que sem internet, conseguem receber o conteúdo das glosas e reproduzi-las no formato de língua de sinais.

A velocidade de reprodução da língua de sinais, geralmente, é inferior a velocidade com que as glosas chegam a aplicação cliente [2]. Por conta disso foi implementada uma fila de glosas a serem reproduzidas, quando o tamanho da fila chega a determinado limiar, a velocidade de reprodução do avatar, no Widget, é acelerada. Caso a aceleração não consiga diminuir o tamanho da fila, ela é truncada, tendo todos seus elementos removidos, forçando uma sincronização para o instante atual, com a penalidade de perder o conteúdo que estava armazenado na fila. Essa abordagem não é a ideal, contudo, é algo que acontece também com intérpretes de língua de sinais humanos, quando estes ficam atrasados em relação ao conteúdo atual que eles estão interpretando [2]. Para funcionamento do módulo, as glosas precisam ser enviadas pela emissora. Atualmente só é possível para vídeos sob demanda

### 4.4 Módulo Áudio descrição

O módulo de áudio descrição requisita ao servidor CCWS o recebimento do áudio recebe como retorno o endereço HTTP para o áudio em formato DASH [9]. Em seguida, é iniciada a reprodução deste áudio. O usuário pode ouvi-lo no viva voz do seu dispositivo móvel ou através de fones de ouvido conectados ao mesmo. Este módulo ainda não foi sincronizado com o conteúdo de legenda e glosas transmitido.

## 5 Principais desafios

Os principais desafios desse trabalho envolveram a sincronização das mídias de legenda, glosa, áudio descrição e vídeo entre os diferentes dispositivos; a disponibilização offline do VLibras Widget [4]; e a Velocidade de exibição da língua de sinais comparada com a velocidade com que as glosas são recebidas.

Soluções foram encontradas para a sincronização entre as mídias de legenda, glosa e vídeo entre vários dispositivos. Contudo, para a mídia de áudio descrição o sincronismo ainda está sendo desenvolvido.

A disponibilização *offline* do Widget e seu dicionário foi implementada e já está funcional. Contudo, ainda estão sendo pesquisadas formas de acelerar a exibição da linguagem de sinais, para manter o conteúdo sincronizado e sem necessidade descartar parte do conteúdo. Uma da possibilidades em análise é sumarização do conteúdo das glosas.

## 6 Conclusões e Trabalhos Futuros

O presente trabalho analisou os requisitos de acessibilidade, no âmbito no projeto TV 3.0, para as mídias de legendas, glosas e áudio descrição. Foram apresentadas soluções para sua implementação, através da proposta de uma extensão da API do CCWS, e de uma arquitetura de sistema, o qual foi implementado como prova de conceito.

Como trabalhos futuros, pretende-se solucionar a sincronização da mídia de áudio descrição com as demais mídias. Também está sendo pesquisado um caminho para manter o sincronismo do avatar de libras com o conteúdo do vídeo, visto que, ele inicia sincronizada, mas depois de um tempo atrasa, devido ao tempo de reprodução da língua de sinais ser maior do que o tempo com que o áudio original é falado. Uma técnica que está sendo analisada é a sumarização da glosa.

Por fim, apesar de se encontrar ainda em fase de desenvolvimento, é possível percebera existência de uma contribuição científica, tecnológica e social da proposta apresentada neste trabalho, uma vez que essa solução, quando implementada na TV 3.0, pode trazer benefícios para aproximadamente 8,8 milhões de deficientes visuais e surdos brasileiros [8].

## References

[1] Technology Committee 24TB. 2013. ST 2052-1:2013 - SMPTE Standard - Timed Text Format (SMPTE-TT). *ST 2052-1:2013* (June 2013), 1–18. https://doi.org/10.5594/SMPTE.ST2052-1.2013

[2] Ursula Bellugi and Susan Fischer. 1972. A comparison of sign language and spoken language. *Cognition* 1, 2 (1972), 173–200. https://doi.org/10.1016/0010-0277(72)90018-2

[3] Advanced Television Systems Committee. 2021. *ATSC Standard: Captions and Subtitles, with Amendments No. 1 and No. 2.* Technical Report. https://prdatsc.wpenginepowered.com/wp-content/uploads/2021/09/A343-2018-Captions-and-Subtitles-with-Amend-1-2.pdf

[4] Universidade Federal da Paraíba (UFPB). 2023. VLibras - Governo Digital. https://vlibras.gov.br/. Online; Accessed on August 30, 2023..

[5] Comissao de Estudo Especial de Televisão Digital (ABNT/CEE-085). 2021. *Standard ABNT NBR 15606-11:2021; Digital Terrestrial Television—Data Coding and Transmission Specification for Digital Broadcasting Part 11: Ginga CC WebServices—Ginga Common Core WebServices Specification.* Technical Report.

[6] Maria Cecília Rafael de Góes. 2020. *Linguagem, surdez e educação.* Autores Associados.

[7] Leslie Haddon and Gerd Paul. 2001. 10. Design in the IT industry: the role of users. *Technology and the market: demand, users and innovation* (2001), 201.

[8] IBGE. 2012. Censo demográfico 2010: características gerais da população, religião e pessoas com deficiência.

[9] British Standards Institution. 2022. *ISO/IEC 23009-1 AMD 1. Information Technology. Dynamic Adaptive Streaming Over HTTP (DASH).: Part 1. Media presentation description and segment formats.* Technical Report pt. 1. https://www.iso.org/standard/83314.html

[10] Pierre-Anthony Lemieux, Nigel Megitt, and Robert Bryer. 2022. imscJS Repository. https://github.com/sandflow/imscJS. Online; Accessed on August 30, 2023..

[11] Sérgio Mattos. 1990. *Um perfil da TV brasileira. 40 anos de história: 1950-1990.* Associação Brasileira de Agências de Propaganda.

[12] Danilo Assis Nobre, Mateus Ferreira, Tiago Maritan U de Araújo, Iris Regina Nascimento, Pollyane Carvalho, and Guido Lemos Filho. 2011. WikiLIBRAS: Collaborative Construction of a Multimedia Dictionary for Brazilian Sign Language. In *Proceedings of the 17th Brazilian Symposium on Multimedia and the Web on Brazilian Symposium on Multimedia and the Web-Volume 1.* 244–251.

[13] TML profiles for Internet media subtitles and captions. 2020. TTML profiles for Internet media subtitles and captions 1.2. https://www.w3.org/TR/ttml-imsc1.2/ Online; Accessed on August 30, 2023..

[14] Luana Silva Reis, Tiago Maritan U de Araújo, Yuska Paola Costa Aguiar, Manuella Aschoff CB Lima, and Angelina S da Silva Sales. 2018. Assessment of the treatment of grammatical aspects of machine translators to Libras. *XXIV Simpósio Brasileiro de Sistemas Multimídia e Web. Anais... Salvador, Brasil: SBC–Sociedade Brasileira de Computação* (2018).

[15] Tatiana Aires Tavares, Celso Alberto Saibel Santos, Thiago Rocha de Assis, Clarissa Braga Bittencourt de Pinho, Germano Mariniello de Carvalho, and Clarissa Santana da Costa. 2007. A TV digital interativa como ferramenta de apoio à educação infantil. *Revista Brasileira de Informática na Educação* 15, 2 (2007), 31–44.

[16] Ângelo Piovesan. 1994. Vídeo e TV na Educação. *Comunicação & Educação* 1 (1994), 105–112.