

Uma Proposta de Protocolo de Codificação de LIBRAS para Sistemas de TV Digital

Felipe Herminio Lemos
Lab. Aplicações de Vídeo Digital
Universidade Federal da Paraíba
João Pessoa - PB
felipel@lavid.ufpb.br

Felipe Lacet Ferreira
Lab. Aplicações de Vídeo Digital
Universidade Federal da Paraíba
João Pessoa - PB
lacet@lavid.ufpb.br

Tiago Maritan U. de Araújo
Lab. Aplicações de Vídeo Digital
Universidade Federal da Paraíba
João Pessoa - PB
maritan@lavid.ufpb.br

Gutenberg Botelho Neto
Lab. Aplicações de Vídeo Digital
Universidade Federal da Paraíba
João Pessoa - PB
gutenberg@lavid.ufpb.br

Guido Lemos Filho
Lab. Aplicações de Vídeo Digital
Universidade Federal da Paraíba
João Pessoa - PB
guido@lavid.ufpb.br

ABSTRACT

Nowadays, the digital television systems do not have specific standards for sign languages (SL). The solution used in TV is to transmit a window with a SL interpreter into the video program. This solution is ineffective in several aspects: it has high operational costs, is dependent of a interpreter, and distract non-deaf viewers. In addition, it does not respects the regional differences in SLs. To reduce these problems, the main objective of this paper is to propose a protocol to encode and transmit Brazilian sign language (LIBRAS) windows into digital television systems. In this propose, the frames of LIBRAS windows are not transmitted. Instead, a set of codes is transmitted, where each code is related to a visual representation of a sign in LIBRAS stored in the TV receiver. Thus it is possible to respect the regional differences of SL and to reduce the used bandwidth. Additionally, some strategies to represent sound effects, voice intonation and emotion nuances are also being investigated.

RESUMO

Os sistemas de TV Digital atuais não possuem padronização específica para língua de sinais (LS). A solução utilizada na TV nos dias de hoje, consiste em transmitir uma janela com um intérprete de LS sobreposto ao vídeo do programa. Esta solução se mostra ineficaz em vários aspectos: possui um alto custo operacional, é dependente de um intérprete em tempo integral, além de distrair os telespectadores que não são surdos. Somado a tudo isso, a geração de uma mesma janela para todos os usuários, não respeita as diferenças regionais das LS. Para reduzir esses problemas, o objetivo principal desse trabalho é propor um protocolo que viabilize a codificação e transmissão de janelas em

Língua Brasileira de Sinais (LIBRAS) para sistemas de televisão digital. Nessa proposta, os quadros do vídeo da janela de LIBRAS não são transmitidos. Em vez disso, é transmitido um conjunto de códigos, onde cada código está relacionado à representação visual de um sinal de LIBRAS armazenada no receptor. Dessa forma, é possível respeitar as diferenças regionais e reduzir a largura de banda utilizada na transmissão. Adicionalmente, também são investigadas estratégias para representar efeitos sonoros, entonação de voz e nuances emocionais na sinalização.

Categories and Subject Descriptors

K.4.2 [Computers and Society]: Social Issues – *Handicapped person/Special Needs*

General Terms

Human Factors, Standardization, Verification

Keyword

LIBRAS, digital television, sign language, encode, protocol, Brazilian Sign Language

1. INTRODUÇÃO

Segundo o censo demográfico do IBGE de 2002, apenas no Brasil, existe cerca de 5,7 milhões de pessoas com algum nível de deficiência auditiva, o que representa cerca de 3,0% da população brasileira [1]. Em termos mundiais, a estimativa da Organização Mundial de Saúde (OMS) é de que aproximadamente 278 milhões de pessoas apresentem algum nível de deficiência auditiva [2].

Para facilitar o acesso à informação através da televisão, um dos recursos de acessibilidade utilizado é o *closed caption* ou legenda oculta. O *closed caption* é um sistema de transmissão de legendas textuais via sinal de televisão. Essas legendas podem ser reproduzidas por um televisor que possua função para tal, e tem como objetivo melhorar o acesso dos deficientes auditivos aos programas transmitidos.

Um das limitações dos principais padrões de transmissão de *closed caption* atuais, no entanto, é a ausência de um suporte claro e padronizado para transmissão de informações associadas às emoções e ao humor dos personagens envolvidos no programa. Esse tipo de informação como, por exemplo, entonação, pausas,

taxa de fala e rouquidão são semanticamente relevantes e sua omissão podem implicar, em alguns casos, em mudança de interpretação e sentido. [3]

Existem alguns trabalhos na literatura que procuram contornar esses problemas, dentre os quais, pode-se destacar o *Emotive Captioning* [4] e o *Emotional Subtitles* [5].

O *Emotive Captioning*, proposto por Rashid *et al.* [4] utiliza tipografia cinética para expressar a emoção em legendas. A tipografia cinética é um conceito que consiste em adicionar animação a um texto para refletir a entonação da voz e emoção. Por exemplo, em uma cena relacionada a medo intenso, o texto da legenda aumentaria rapidamente de tamanho e vibraria rapidamente. Cores e imagens também são utilizados para representar emoção (ver Figura 1). Uma das limitações dessa solução é a falta de avaliação por usuários.



Figura 1: Legenda emotiva dinâmica, para representar engarrafamento no trânsito, extraída de [4].

O *Emotional Subtitles*, proposto por Ohene-Djan *et al.* [5], é um sistema em que a semântica da sentença é definida pelo usuário. Nesse sistema, é possível descrever as características emocionais associadas com uma frase ou um trecho do texto. Essas características podem ser representadas através de atributos como volume, cor, tamanho, texto e entonação. Os possíveis valores de entonação são os seguintes: feliz, triste, sarcástico, animado, cômico, suplicante, questionador, com medo, com autoridade e com raiva.

Para facilitar a criação de vídeos legendados com representação de emoções, Ohene-Djan *et al.* também desenvolveram um editor de legendas emocionais. Este editor permite que os usuários entrem com a legenda e formate-as de acordo com esquemas pré-definidos, mudando apenas a fonte e a cor do texto. Os usuários também podem alterar o tamanho do texto de acordo com a entonação de voz do interlocutor, conforme ilustrado na Figura 2. Outra característica interessante, é que, para auxiliar na identificação do personagem que está discursando, a cor da legenda coincide com a cor do rosto do personagem.

No entanto, para os surdos, as línguas orais, utilizadas nas legendas ocultas, representam apenas “uma segunda língua” [6]. A forma natural de comunicação entre eles é através de línguas de sinais. As línguas de sinais são línguas naturais de modalidade gestual-visual, em que gestual significa o conjunto de elementos lingüísticos manuais, corporais e faciais necessários para a articulação e significação visual-cultural do sinal [7].

Cada comunidade de surdos desenvolveu a sua língua de sinais ao longo dos tempos, assim como cada comunidade de ouvintes desenvolveu a sua língua oral. Por esta razão, em geral, cada país possui sua própria língua de sinais.

No Brasil, por exemplo, a língua de sinais utilizada pela maioria dos surdos brasileiros e reconhecida pela lei nº 10.436, de 24 de abril de 2002 é a língua brasileira de sinais - LIBRAS [8].

Os sistemas de TV Digital atuais, no entanto, não possuem padronização específica para língua de sinais. A solução utilizada na TV nos dias de hoje, consiste em transmitir uma janela com um intérprete de LIBRAS sobreposto ao vídeo do programa. Esta solução se mostra ineficaz na TV digital em vários aspectos: (1) custo operacional, pois requer uma infraestrutura para geração e produção; (2) é dependente de um intérprete em tempo integral, (3) além de distrair os telespectadores que não são surdos. Além disso, essa solução não oferece suporte às diferenças regionais das línguas de sinais.



Figura 2: O programa de televisão BBC's Tweenies com legenda emocional, extraída de [5].

No Brasil, por exemplo, alguns sinais de LIBRAS no sudeste do país são representados de forma diferente dos sinais correspondentes no nordeste do Brasil. Em virtude disso, a geração de uma mesma janela de LIBRAS para todos os usuários, não respeita essas diferenças regionais e pode dificultar a compreensão.

Para minimizar esses problemas, o objetivo principal desse trabalho é propor um protocolo que viabilize a codificação e transmissão da janela de LIBRAS em sistemas de TV Digital, incluindo maneiras de representar efeitos sonoros, entonação de voz e nuances emocionais na sinalização.

Nessa proposta de codificação, os quadros do vídeo da janela de LIBRAS não são transmitidos. Em vez disso, um conjunto de códigos é transmitido, onde cada código está associado à representação visual de um sinal de LIBRAS que está armazenada no receptor, conforme proposto por Araújo *et al.* [11] (ver Seção 3). Dessa forma, é possível respeitar as diferenças regionais de LIBRAS, além de reduzir a largura de banda utilizada para transmitir a janela de LIBRAS.

As demais seções do artigo estão organizadas da seguinte forma. Na Seção 2 são revisados os principais conceitos da Língua Brasileira de Sinais. Na seção 3 é apresentado o sistema de geração automática de janelas de LIBRAS para sistemas de TV digital, proposto por Araújo *et al.* [11]. Na Seção 4 é apresentada a proposta de protocolo para codificação de LIBRAS. Na Seção 5 são apresentados alguns experimentos aplicados para validar a proposta. Por fim, as considerações finais são apresentadas na Seção 6.

2. LÍNGUA BRASILEIRA DE SINAIS

A LIBRAS é uma língua natural surgida entre os deficientes auditivos brasileiros. Brito (1995) [9] afirma que é uma língua natural porque, como as línguas orais, LIBRAS surgiu espontaneamente da interação entre os deficientes auditivos, além de poderem expressar qualquer conceito desde o descritivo/concreto ao emocional/abstrato.

O equivalente da palavra (ou item lexical) nas línguas oral-auditivas é denominado de sinal nas línguas de sinais. Os sinais são formados por meio da combinação de formas e de movimentos das mãos e de pontos de referência no corpo ou no espaço. Eles são compostos pelos seguintes parâmetros [10]:

- configuração das mãos: são as formas das mãos que podem ser da datilografia (alfabeto manual) ou outras formas feitas pela mão predominante (mão direita para os destros ou esquerda para os canhotos), ou pelas duas mãos.
- ponto de articulação: é o lugar onde incide a mão predominante configurada, ou seja, local onde é feito o sinal, podendo tocar alguma parte do corpo ou estar em um espaço neutro.
- movimento: os sinais podem ter um movimento ou não. Os movimentos podem ser internos às mãos (movimentos dos dedos) ou deslocamentos das mãos em relação ao corpo.
- expressão facial e/ou corporal: as expressões faciais / corporais são de fundamental importância para o entendimento real do sinal, sendo que a entonação em língua de sinais é feita pela expressão facial.
- orientação/direção: Os sinais têm uma direção com relação aos parâmetros acima. Assim, os verbos IR e VIR se opõem em relação à direcionalidade.

Para se comunicar em LIBRAS, no entanto, não basta apenas conhecer os sinais isoladamente, é necessário também conhecer a estrutura gramatical da língua.

3. SISTEMA DE GERAÇÃO AUTOMÁTICA DE JANELAS DE LIBRAS

Nesta seção será apresentada brevemente a solução proposta por Araújo *et al.* [11] para geração automática de janelas de LIBRAS para sistemas de TV digital. A principal contribuição desta solução é a utilização de um dicionário de LIBRAS no receptor. O dicionário de LIBRAS é responsável pelo armazenamento das representações visuais dos sinais de LIBRAS. Nesse dicionário, cada sinal (ou expressão) pode ser representado por uma animação ou um arquivo de vídeo e possui um código (ou índice) associado a essa representação.

Na Figura 3 é apresentada a arquitetura da solução proposta por Araújo *et al.* [11]. Conforme pode ser observado nessa figura, inicialmente, na estação de TV, um componente Gerador de LIBRAS recebe um fluxo de entrada de áudio ou *closed caption* de um codificador ou de um extrator de *closed caption*, respectivamente. Em seguida, utilizando uma Gramática de LIBRAS, o Gerador de LIBRAS gera automaticamente um fluxo elementar (*elementary stream – ES*) em língua de sinais. Esse ES é composto por um conjunto de códigos, onde cada código está relacionado a uma animação armazenada em um Dicionário de

LIBRAS. Após ser gerado, esse ES é multiplexado (ou remultiplexado) no fluxo MPEG-2 *Transport Stream* (MPEG-2 TS) da estação de TV e transmitido no sinal de TV.

Do outro lado, no *set-top-box*, existe um dispositivo de memória estendida, como, por exemplo, um dispositivo de armazenamento USB, que armazena o Dicionário de LIBRAS e uma aplicação nativa que utiliza esse dicionário, para decodificar e apresentar os sinais na tela.

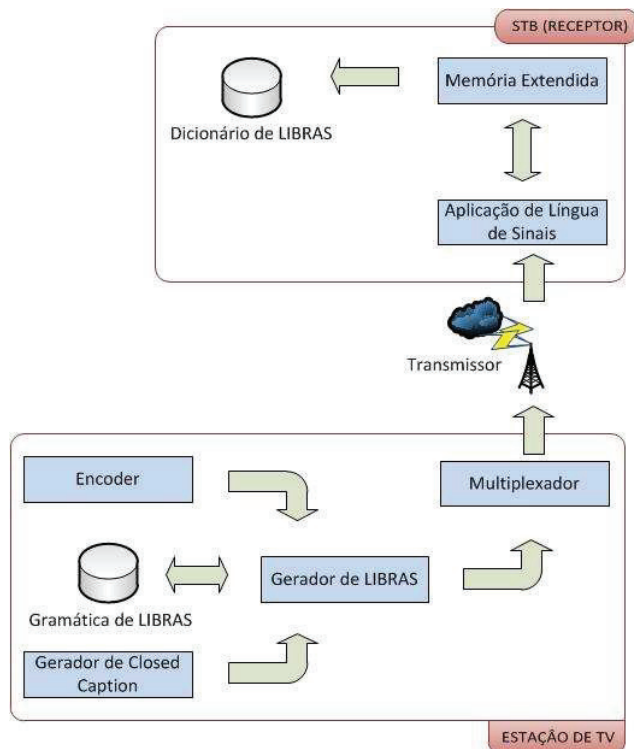


Figura 3: Arquitetura proposta para a transmissão de LIBRAS no Sistema Brasileiro de TV Digital.

Uma característica importante da solução proposta por Araújo *et al.* [11] é que, como apenas um conjunto de códigos é transmitido e os vídeos da janela são montados no receptor de TV com base em um Dicionário de LIBRAS, os aspectos regionais da LIBRAS podem ser respeitados. Inspirado nessa idéia, o protocolo de codificação proposto nesse trabalho, codifica em suas mensagens esse conjunto de códigos, em vez dos quadros do vídeo da janela de LIBRAS.

4. PROTOCOLO PROPOSTO

Nessa seção apresentaremos a proposta de protocolo de codificação de janelas de LIBRAS, proposto nesse trabalho. Nessa proposta de codificação, os quadros do vídeo da janela de LIBRAS não são transmitidos. Em vez disso, um conjunto de códigos é transmitido, onde cada código está associado à representação visual de um sinal de LIBRAS, armazenada no dicionário de LIBRAS do receptor (ver Seção 3). Conforme mencionado na Seção 1, a principal motivação para codificar as janelas de LIBRAS dessa forma é permitir que as diferenças regionais (regionalismos) da LIBRAS sejam respeitadas, além de reduzir a largura de banda utilizada para transmitir a janela de LIBRAS.

Dessa forma, cada usuário poderia utilizar um dicionário de LIBRAS com as representações dos sinais de sua região, adaptando a apresentação da janela de LIBRAS ao seu contexto. Além disso, o protocolo também permite que informações associadas às emoções, ao humor e alguns efeitos sonoros sejam codificadas e transmitidas.

O protocolo define dois tipos de mensagens: uma mensagem de controle, a *librasControlMessage* (LCM), e uma mensagem de dados, a *librasDataMessage* (LDM). As mensagens LCM transmitem informações de controle sobre a janela de LIBRAS como, por exemplo, a resolução da camada gráfica de apresentação, o tamanho e a posição inicial da janela, etc. As mensagens LDM, por outro lado, transmitem o conjunto de códigos dos sinais de LIBRAS e as informações associadas às emoções, ao humor e aos efeitos sonoros como, por exemplo, a, entonação de voz, algumas nuances emocionais, etc. Além disso, na LDM também é possível identificar o personagem que está discursando.

A sintaxe das mensagens LCM e LDM é apresentada nas Tabelas 1 e 2, respectivamente. Conforme ilustrado nessas tabelas, as duas mensagens sempre iniciam com um campo **message_id**, que identifica o tipo da mensagem (ver Tabela 4).

Conforme pode ser observado na Tabela 1, a LCM é composta pelos seguintes campos: resolução, posição_x, posição_y, largura e altura. O campo resolução define a resolução da camada gráfica de apresentação da janela de LIBRAS, como, por exemplo, 1920x1080, 720x480, etc. Os campos posição_x e posição_y definem a posição inicial da janela, isto é, as coordenadas x e y, respectivamente, tomando como referência, (i.e., como coordenada (0,0)) a borda superior esquerda. Os campos largura e altura definem a largura e a altura inicial da janela, respectivamente.

A mensagem LDM (ver Tabela 2) é composta principalmente pelos campos: *identifier_data_bytes*, *parameters_data_bytes* e *glosa_data_bytes*.

Tabela 1: Estrutura da *LibrasControlMessage*.

librasControlMessage {	
message_id	8 bits
message_length	16 bits
resolução	8 bits
posição_x	16 bits
posição_y	16 bits
largura	16 bits
altura	16 bits
private_data_length	8 bits
for (i=0; i < N; i++) {	
private_data_bytes;	8 bits
}	
}	

O campo *identifier_data_bytes* é utilizado para identificar o personagem que está falando. Por exemplo, se o personagem Horacio está discursando na cena, o valor HORACIO é representado no campo *identifier_data_bytes*, de forma similar ao formato utilizado para identificar o personagem em legendas ocultas (ver Figura 5).

Efeitos sonoros e as emoções transmitidas pelos personagens são armazenados no campo *parameters_data_bytes*.

Na Tabela 3 são apresentados alguns possíveis valores para esse campo.

Por fim, o campo *glosa_data_bytes* é utilizado para representar os conjuntos de códigos que representam os sinais de LIBRAS. Neste trabalho, o código de cada sinal será simplesmente a *glosa* (representação textual) daquele sinal. Por exemplo, o código do sinal que representa a palavra casa em LIBRAS é a *glosa* “CASA”.

Cada mensagem LDM pode representar um sinal único ou uma sentença completa (conjunto de sinais). As palavras (ou expressões) que não possuem um sinal correspondente em LIBRAS, como, por exemplo, os nomes próprios, são codificados de forma soletrada, i.e., codificadas como um conjunto de glosas, onde cada glosa representa uma letra do nome.

Tabela 2: Estrutura da *LibrasDataMessage*.

librasDataMessage {	
message_id	8 bits;
message_length	16 bits;
reservado	7 bits;
identifier_control_flag	1 bit;
if (identifier_control_flag = 1) {	
identifier_data_length	8 bits;
for(i=0; i < N, i++) {	
identifier_data_bytes	8 bits;
}	
}	
reservado	7 bits;
parameters_control_flag	1 bit;
if (parameters_control_flag = 1) {	
parameters_loop_length	8 bits;
for(i=0; i < N, i++) {	
parameters_data_length	8 bits;
for(i=0; i < N, i++) {	
parameters_data_bytes	8 bits;
}	
}	
else {	
sinais_loop_length	16 bits;
for(i=0; i < N, i++) {	
glosa_data_length	8 bits;
for(i=0; i < N, i++) {	
glosa_data_bytes	8 bits;
}	
}	
}	

4.1. Sincronização e Encapsulamento

Outro requisito importante do protocolo de codificação é definir qual a estrutura usada para encapsular as mensagens do protocolo em um fluxo de TV digital, MPEG-2 *Transport Stream* (MPEG-2 TS) e como apresentar os sinais sincronizadamente com o vídeo do programa.

Uma possível solução para os dois problemas seria transportar as mensagens no protocolo de codificação de eventos definido pela especificação *Digital Storage Media-Command and Control* [12], os eventos DSM-CC.

Os eventos DSM-CC, transmitidos em estruturas denominadas *Stream Event Descriptor* (ver Tabela 5), permitem que pontos de sincronização sejam definidos em nível de aplicação. Nesse tipo de estrutura, o campo *eventNPT* carrega uma etiqueta de tempo (*timestamp*) relacionada ao relógio de referência do fluxo MPEG-2 TS. Isto torna possível que os sinais sejam sincronizados com outras mídias, como o vídeo ou o áudio do programa. Essa estrutura também define um espaço para transporte de dados privados (campo *privateDataPayload*) que, nesse caso, poderia ser utilizada para transportar as mensagens do LCM e LDM do protocolo proposto.

Desta forma, uma aplicação que recebesse esses eventos poderia extrair as mensagens do protocolo, decodificar as informações transportadas na mensagem e exibir os sinais sincronizadamente com o vídeo do programa.

5. EXPERIMENTOS

Nessa seção descrevemos os experimentos utilizados para validar a solução proposta. Inicialmente, apresentamos como a solução foi implementada no Sistema Brasileiro de Televisão Digital (SBTVD). Em seguida, apresentamos alguns experimentos realizados com usuários surdos e intérpretes de LIBRAS.

5.1. Implementação para o SBTVD

Inicialmente, foi desenvolvida utilizando a linguagem de programação C++, uma ferramenta responsável por codificar os sinais de LIBRAS de acordo com o protocolo apresentado na Seção 4, denominado, Codificador de LIBRAS. Nessa implementação, um usuário digita textos em *glossa* (i.e., na estrutura gramatical de LIBRAS) que são codificados pela ferramenta, encapsulados em eventos DSM-CC e empacotados em pacotes MPEG-2 *Transport Stream* (MPEG-2 TS) para em seguida serem multiplexados no fluxo MPEG-2 TS da TV.

Tabela 3: Possíveis valores para o *parameters_data_byte*.

Emoções / Efeitos Sonoros	Descrição
[GRITOS]	Demonstra som de gritos
[RISOS]	Demonstra som de risada
[TIRO]	Demonstra som de tiro
[Maria chorando]	Demonstra que a personagem Maria está chorando. "Maria" é o identificador do personagem e é codificado no campo <i>identify_data_bytes</i> da LDM. A expressão "chorando" representa a emoção do personagem e é codificado no campo <i>parameters_data_byte</i> da LDM.
[SINOS]	Demonstra som de sinos
♪ ♪	Demonstra uma música em segundo plano, ou que algum personagem está cantando.

Para validar o fluxo gerado por essa ferramenta, do lado do receptor foi desenvolvida uma aplicação interativa Ginga-J [13] utilizando o OpenGinga, uma implementação de código aberto do middleware Ginga [13], disponível em [14].

Para validar o fluxo gerado por essa ferramenta, do lado do receptor foi desenvolvida uma aplicação interativa Ginga-J [13]

utilizando o OpenGinga, uma implementação de código aberto do middleware Ginga [13], disponível em [14].

Tabela 4: Valores possíveis para o campo *message_id*.

message_id	message type LIBRAS
0x01	librasControlMessage
0x02	librasDataMessage
0x03-0xFF	reserved for future use

Essa aplicação extrai as informações de LIBRAS do fluxo e, com o auxílio de um dicionário de LIBRAS (ver Seção 3), decodifica, sincroniza e apresenta os sinais de LIBRAS. O dicionário de LIBRAS utilizado nessa implementação, representa os sinais através de um agente animado virtual 3D e foi desenvolvido utilizando o software Blender, disponível em [15].



Figura 5: Cena do desenho 101 Dálmatas com *closed caption*, demonstrando a identificação do personagem na legenda oculta.

Tabela 5: Estrutura do *Stream Event Descriptor*.

```

StreamEventDescriptor {
    descriptorTag          8 bits
    descriptorLength      8 bits
    eventId                8 bits
    Reserved               31 bits
    eventNPT               33 bits
    privateDataLength     8 bits
    commandTag            8 bits
    sequenceNumber        7 bits
    finalFlag             1 bit
    privateDataPayload    8 a 2008 bits
    FCS                   8 bits
}

```

A Figura 6 apresenta uma captura de tela da janela de LIBRAS gerada pela aplicação interativa, a partir da codificação gerada pelo Codificador de LIBRAS.

Para validar a codificação dos efeitos sonoros, das emoções e da identificação do personagem, na próxima subseção apresentamos como essas informações são representadas para o usuário, i.e., como a aplicação interativa deve decodificar e apresentar essas informações.

Para validar a codificação dos efeitos sonoros, das emoções e da identificação do personagem, na próxima subseção apresentamos

como essas informações são representadas para o usuário, i.e., como a aplicação interativa deve decodificar e apresentar essas informações.

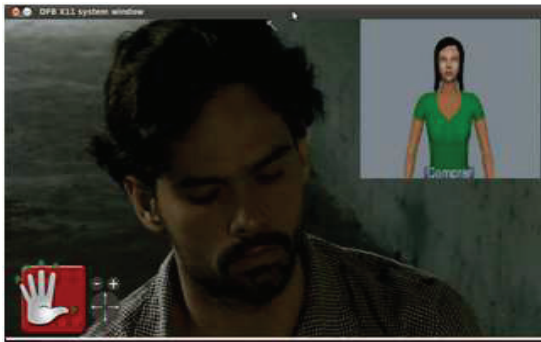


Figura 6: App LIBRAS-TV.

5.1.1 Representação de efeitos sonoros, emoções e identificação dos personagens

Os efeitos sonoros extraídos das mensagens LDM são apresentados aos usuários através de pistas visuais. As pistas visuais são imagens que descrevem iconicamente o efeito sonoro na cena atual, ou em segundo plano, conforme ilustrado na Tabela 6. Essas pistas visuais são apresentadas no canto superior da tela juntamente com a janela de LIBRAS. Dessa forma, os usuários podem interpretar o efeito sonoro tanto pela imagem quanto pela cor da elipse, proporcionando uma melhor compreensão do conteúdo da cena.

As emoções geralmente são representadas através da expressão facial, mas dependendo do contexto, ela deve ser representada com mais ênfase. Para associarmos a emoção ao personagem como também identificar o falante da cena, utilizamos uma abordagem similar a de Ohene-Djan J. *et al.* [5], onde imagens com cores e números identificam quem está falando. Além disso, essa imagem também identifica quem está sentindo a emoção, combinando a representação da emoção realizada pelo intérprete de LIBRAS (agente animado virtual) com a imagem associada ao personagem. As imagens utilizadas inicialmente são *emotions* com cores diferentes acompanhados do alfabeto *alfa-numérico*, conforme ilustrado na Figura 11.

Outra forma de identificar os personagens de um diálogo é através da inclinação do intérprete (ou agente animado virtual). Nesse caso, quando um personagem discursa, o intérprete se inclina em um sentido (para a esquerda, por exemplo) e quando outro personagem discursa, o intérprete se inclina em outro sentido (para a direita, por exemplo). As Figuras 7, 8, 9 e 10 ilustram a representação de efeitos sonoros, de emoções e da identificação de personagens geradas pela aplicação interativa.

5.2 Experimentos Com Usuários

Após implementar a solução, alguns testes com usuários surdos e intérpretes de LIBRAS foram aplicados para avaliar a viabilidade e adequabilidade da solução.

Conforme sugerido por Eronen [2006] [16], entrevistas usando métodos de observação são importantes em estudos exploratórios, pois fornecem dados sobre o comportamento do visualizador e possíveis explicações para esse comportamento. Estamos tentando estudar as reações do espectador a um novo conceito, LIBRAS

com legendas reforçadas, e pretendemos captar as reações junto com o raciocínio por trás delas em tempo real.

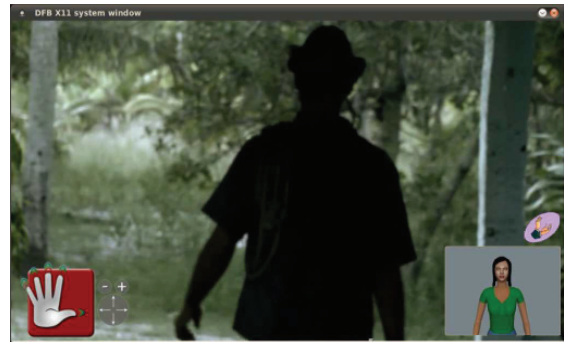


Figura 7: Demonstração do efeito sonoro (aplausos).

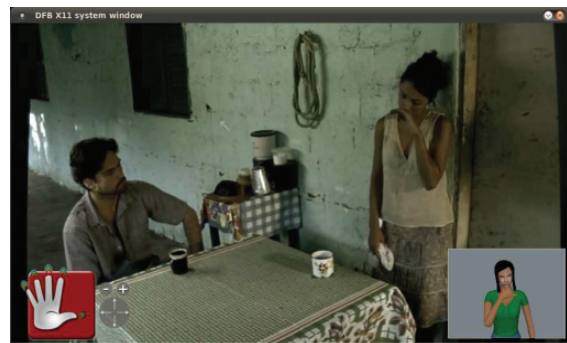


Figura 8: Demonstração de emoção sentida pelo personagem.

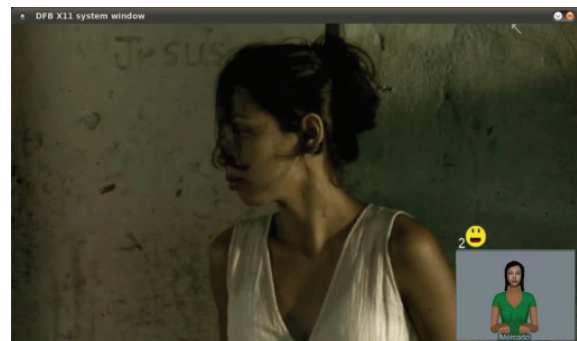


Figura 9: Identificação do personagem através de imagens.

Os testes foram realizados com doze surdos e dois intérpretes de LIBRAS de João Pessoa/Paraíba. O grupo foi formado por homens e mulheres, entre 16 e 37 anos, e com variados grau de instrução.



Figura 11: Imagens utilizadas para identificação do personagem.





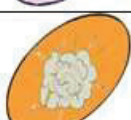

Os usuários foram convidados a assistir cinco vídeos e responder a um questionário com o objetivo de avaliar os seguintes aspectos: diferenças regionais, uso correto da LIBRAS, identificação dos personagens, representação da emoção e representação de efeitos sonoros.



Figura 10: Identificação do personagem através da inclinação do avatar.

A descrição dos vídeos usados nos testes são apresentados na Tabela 7. Um intérprete de LIBRAS estava presente durante os testes para auxiliar os surdos no preenchimento do questionário de avaliação.

Tabela 6: Pistas visuais para sons do ambiente.

Pista Visual	Significado
	Música de fundo
	Tiro
	Trovão
	Gargalhada
	Explosão
	Música sendo cantada por personagem

Na Figura 12, são apresentados os resultados da avaliação das diferenças regionais. Nesse teste, os usuários (todos naturais de João Pessoa) foram divididos em dois grupos aleatoriamente. Um grupo assistiu e avaliou o vídeo com os sinais da região de São Paulo e o outro assistiu e avaliou o vídeo com os sinais da região de João Pessoa. O eixo 'X' do gráfico representa o número de surdos que optaram pelas alternativas existentes no eixo 'Y'.

Conforme pode ser observado na Figura 12, pode-se perceber que as diferenças regionais influenciam bastante na compreensão do conteúdo representado. Os usuários que assistiram ao vídeo com

os sinais de João Pessoa compreenderam melhor que os usuários que assistiram o vídeo com os sinais de São Paulo.

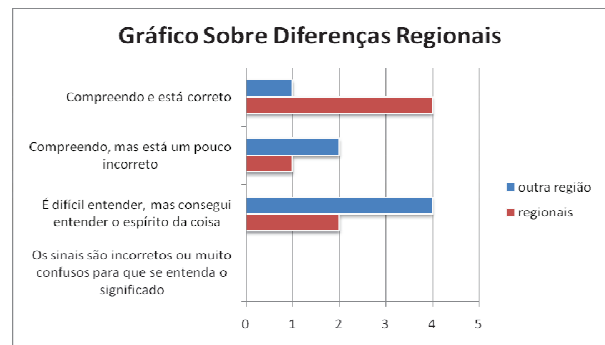


Figura 12: Gráfico sobre as diferenças regionais.

Tabela 7: Tabela dos vídeos usados para testes

Vídeo	Objetivo	Duração
Cena de um filme com uma janela de LIBRAS com sinais de João Pessoa.	Avaliar o nível de entendimento dos surdos com sinais de sua região	65 seg.
Cena do mesmo filme com uma janela de LIBRAS com sinais de São Paulo.	Avaliar o nível de entendimento dos surdos com sinais de diferentes regiões	65 seg.
Cena de um filme com identificação do personagem através da inclinação do avatar.	Avaliar o entendimento dos surdos sobre quem está falando na cena	65 seg.
Cena de um filme com identificação do personagem através de imagens.	Avaliar o entendimento dos surdos sobre quem está falando na cena	65 seg.
Dois frases representando emoções	Avaliar o nível de entendimento dos surdos sobre as emoções sentidas pelos personagens	10 seg.
Uma frase onde demonstrava-se efeitos sonoros	Avaliar o nível de entendimento dos surdos sobre as pistas visuais que representam efeitos sonoros	10 seg.

Na Figura 13, são apresentados os resultados obtidos com relação à identificação dos personagens. Nesse teste, os usuários assistiram e avaliaram as duas estratégias para identificação do personagem, através de imagens (emoticons) ou através da inclinação do agente animado virtual. Conforme pode ser observado na Figura 13, um maior número de usuários conseguiu identificar o personagem através do uso de imagens. As pistas visuais não tiveram um resultado satisfatório, os usuários não entenderam ou confundiram o entendimento sobre o que a frase estava querendo demonstrar.

Nas Figuras 14 e 15, os usuários avaliaram a capacidade de uso do sistema na ausência de intérprete humano e o grau de relevância

da solução proposta, a Figura 14 mostra que os usuários reconhecem a relevância do sistema, porém boa parte dos usuários ainda considera impossível utilizar o sistema na ausência de um intérprete humano por não ter o hábito de ver a LIBRAS realizada por um avatar 3D, como mostra a Figura 15.

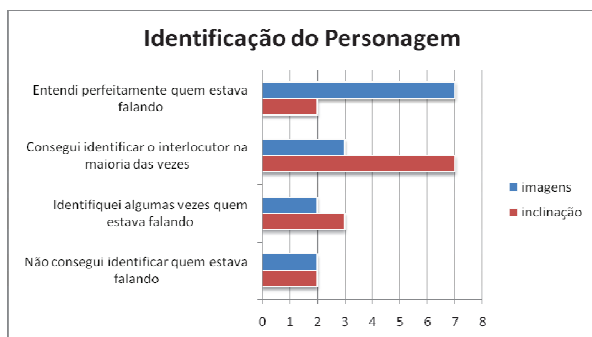


Figura 13: Gráfico sobre a identificação do personagem.

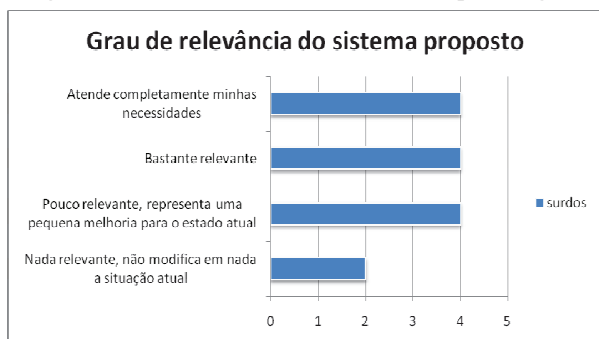


Figura 14: Gráfico que demonstra a opinião dos surdos sobre o grau de relevância do sistema proposto.

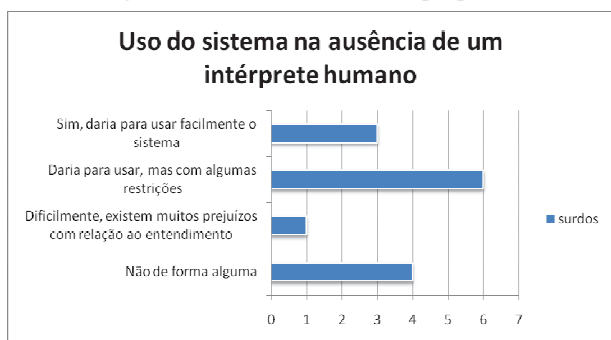


Figura 15: Gráfico que demonstra a aprovação dos surdos para o uso do sistema na ausência de um intérprete humano.

6. CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma proposta de protocolo de codificação de janelas de LIBRAS para Sistemas de TV Digital. A solução proposta permite que as diferenças regionais presentes em LIBRAS sejam respeitadas, além de permitir que informações relacionadas a efeitos sonoros, emoções, e a identificação do personagem que está discursando também sejam transmitidas.

Alguns testes foram aplicados com surdos para validar a solução proposta. Nesses testes, ficou evidente que as diferenças regionais afetam a compreensão e, conseqüentemente, a comunicação. Em virtude disso, as soluções tradicionais que transmitem uma janela

única para usuários de diferentes regiões não atende às necessidades dos usuários.

Dentre as propostas de trabalho futuro, está a de validar a solução em sistemas que renderizam os sinais de LIBRAS em tempo real e desenvolver outras estratégias para identificação dos personagens e representação de efeitos sonoros.

7. REFERÊNCIAS

- [1] GOTTI, M. O. Políticas, métodos e técnicas de desenvolvimento do bilingüismo dos surdos. Disponível em: <http://dtit.unilat.org/tercer_seminario/actas/oliveira_gotti_pt.htm>. Acesso em: 17 Abril 2011.
- [2] WHO - Deafness and hearing impairment. World Health Organization, 2010. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs300/en/>>. Acesso em: 17 Abril 2011.
- [3] JUSLIN, P. N. AND LAUKKA, P. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin* v. 129, n. 5, p. 770-814.
- [4] Lee, D.G., Fels, D.I., and Udo, J.P. (2007). Emotive Captioning. *ACM Comput, Entertain.* v. 5, n. 2, p. 3-15 Article 11 (August 2007).
- [5] OHENE-DJAN, J.; WRIGHT, J; COMBIE-SMITH, K. Emotional Subtitles: A System and Potential Applications for Deaf and Hearing Impaired People. Goldsmiths College, University of London, New Cross, London, p. 2-5, 2007.
- [6] GOES, M. C. R.. Linguagem, Surdez e Educação. Campinas: Editora Autores Associados, 1996.
- [7] CAMPOS, M. B.; GIRAFÁ, L. M. M. SIGNSIM: uma ferramenta para auxílio à aprendizagem da língua brasileira de sinais. V Congresso Ibero-Americano de Informática na Educação – RIBIE. Viñadelmar, Chile: [s.n.]. 2000. p. 1-13.
- [8] FUSCO, E.. X-LIBRAS: Um ambiente virtual para a língua brasileira de sinais. Centro Universitário "Eurípedes de Marília" - UNIVEM. Marília, p. 158. 2004.
- [9] BRITO, L. F.. Por uma gramática de língua de sinais. Rio de Janeiro: Editora Tempo Brasileiro, 1995.
- [10] DENARDI, R. M.; MENEZES P. B.; COSTA A. C. R. AGA-Sing: Animador de Gestos Aplicado à língua de sinais. INFOCOMP (UFLA), Lavras, Minas Gerais, Brasil, v. 4, n. 1, p. 47-53, 2005.
- [11] ARAUJO, T. M. U.; SOUZA, G. L. F.; et AL (2009), p. 1-2. An architecture to generate automatic Brazilian sign language legends into Digital Television Systems. 7th European Conference on Interactive TV.
- [12] ISO/IEC 13818-6. International Organization for Standardization/International Electrotechnical Commission, "Information Technology - Generic coding of moving pictures and associated information: Part 6: Extension for Digital Storage Media Command and Contr. [S.1.]. 1996.
- [13] ABNT NBR 15606-4. Televisão digital terrestre - Codificação de dados e especificações de transmissão para transmissão digital - Parte 4: Ginga-J - Ambiente para a execução de aplicações procedurais. 2010.
- [14] GingaCDN - Code Development Network. Disponível em: <<http://gingacd.n.lavid.ufpb.br/>>. Acesso em: 30 Abril 2011.
- [15] Blender, 2011. Disponível em: <<http://www.blender.org/>>. Acesso em: 17 Abril 2011.
- [16] ERONEN, L. 2006. Five qualitative research methods to make iTV applications universally accessible. *Universal Access in Information Society* v. 5, n. 3, p. 219-238.