

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

UMA PROPOSTA DE PROTOCOLO DE
CODIFICAÇÃO DE LIBRAS PARA
SISTEMAS DE TV DIGITAL

FELIPE HERMINIO LEMOS

JOÃO PESSOA-PB

Novembro-2012

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

**UMA PROPOSTA DE PROTOCOLO DE CODIFICAÇÃO
DE LIBRAS PARA SISTEMAS DE TV DIGITAL**

FELIPE HERMINIO LEMOS

JOÃO PESSOA-PB

Novembro-2012

FELIPE HERMINIO LEMOS

**UMA PROPOSTA DE PROTOCOLO DE CODIFICAÇÃO
DE LIBRAS PARA SISTEMAS DE TV DIGITAL**

DISSERTAÇÃO APRESENTADA AO CENTRO DE INFORMÁTICA DA
UNIVERSIDADE FEDERAL DA PARAÍBA, COMO REQUISITO PARCIAL
PARA OBTENÇÃO DO TÍTULO DE MESTRE EM INFORMÁTICA.

ÁREA DE CONCENTRAÇÃO: SISTEMAS DIGITAIS (TELEVISÃO
DIGITAL)

Orientador: Prof. Dr. Guido Lemos de Souza Filho

JOÃO PESSOA-PB

Novembro-2012

SUMÁRIO

SUMÁRIO.....	iii
Índice de Figuras	vii
Índice de Tabelas.....	ix
Lista de Acrônimos e Siglas	x
AGRADECIMENTOS	xi
RESUMO	xii
ABSTRACT	xiii
1. INTRODUÇÃO.....	14
1.1. Motivação	16
1.2. Objetivos.....	19
1.3. Escopo do Trabalho	19
1.4. Organização do Trabalho	20
2. LIBRAS NO SISTEMA BRASILEIRO DE TV DIGITAL.....	22
2.1. TV Digital	22
2.1.1. Sistemas de TV digital.....	23
2.2. Eventos de Sincronismo DSM-CC	25
2.3. Sistema Brasileiro de TV Digital (SBTVD)	27
2.4. Língua Brasileira de Sinais	28
2.4.1. Espaço de sinalização	32
2.5. Considerações Finais	33
3. TRABALHOS RELACIONADOS	34
3.1. Representação de Emoções e Sons Ambientes para Deficientes Auditivos	34
3.1.1. <i>Emotive Captioning</i>	34
3.1.2. <i>Emotional Subtitles</i>	36
3.1.3. <i>Dynamic Captioning</i>	37
3.2. Transmissão de Línguas de Sinais	39
3.2.1. <i>MobileASL</i>	39
3.2.2. Prioridade na codificação da região de interesse para língua de sinais	40
3.2.3. Sistema de codificação de vídeo para comunicação da língua de sinais a baixas taxas de bits	41

3.2.4. <i>Rybená</i>	42
3.3. Considerações Finais	43
4. PROJETOS DE ACESSIBILIDADE	45
4.1. Projeto LibrasTV	45
4.1.1. Gerador de LIBRAS	47
4.1.2. Exibidor	51
4.1.3. LibrasTV em sistemas digitais	52
4.2. Grupo de Trabalho Acessibilidade como um Serviço (GTAAaS).....	53
4.3. CineLIBRAS.....	55
4.4. Considerações Finais	58
5. PROTOCOLO PROPOSTO.....	59
5.1. Sincronização e Encapsulamento.....	68
5.2. Implementação para o SBTVD.....	69
5.2.1. Representação de efeitos sonoros, emoções e identificação dos personagens	70
5.3. Considerações Finais	74
6. RESULTADOS E DISCUSSÕES.....	76
6.1. Ambiente de Testes.....	76
6.2. Testes Computacionais Objetivos.....	77
6.3. Testes com os Usuários.....	81
6.3.1. Instrumentação.....	82
6.3.2. Primeiro grupo de testes	84
6.3.3. Segundo grupo de testes	89
6.4. Considerações Finais	95
7. CONCLUSÃO E CONSIDERAÇÕES FINAIS	97
7.1. Resultados Alcançados	97
7.2. Trabalhos Futuros e Sugestões	97
7.3. Contribuições	98
REFERÊNCIAS BIBLIOGRÁFICAS	99
ANEXO A – QUESTIONÁRIOS DE AVALIAÇÃO	103
ANEXO B – PRODUÇÃO TÉCNICA E CIENTÍFICA	111

Índice de Figuras

Figura 1: Exemplos de variações regionais nas línguas orais.	18
Figura 2: Exemplos de variações regionais nas línguas de sinais, extraída de [20].	18
Figura 3: Arquitetura de um ambiente de transmissão de TV Digital.	23
Figura 4: Exemplo de configuração de mão em sinais, extraída de [20].	29
Figura 5: Configurações de mão da LIBRAS, extraída de [20].	29
Figura 6: Exemplo de pontos de articulação, extraída de [20].	30
Figura 7: Exemplo de movimento, extraída de [20].	30
Figura 8: Exemplo de expressões faciais, extraída de [20].	31
Figura 9: Exemplo da orientação dos sinais, extraída de [20].	31
Figura 10: Diagrama do espaço de sinalização.	32
Figura 11: Legenda emotiva dinâmica, para representar engarrafamento no trânsito, extraída de [23].	35
Figura 12: O programa de televisão BBC's Tweenies com legenda emocional, extraída de [24].	36
Figura 13: Exemplos de diferentes estilos de legendas: a) legendas scroll-up, b) legendas pop-up, c) legendas pain-on, d) legendas cinematográficas, e) legendas dinâmicas, extraído de [25].	37
Figura 14: Exemplo de legenda dinâmica, extraída de [25].	38
Figura 15: Ilustração esquemática da proposta de Richang Hong et al., extraído de [25].	38
Figura 16: Screenshot do codec MobileASL sobre o celular HTC TyTN II, extraído de [28].	40
Figura 17: Região de interesse para codificação, extraído de [29].	41
Figura 18: Resultados da codificação com o método proposto, extraído de [30].	42
Figura 19: Torpedo Rybená [31].	43
Figura 20: Arquitetura proposta para a transmissão de LIBRAS no Sistema Brasileiro de TV Digital.	46
Figura 21: Detalhamento do Gerador de LIBRAS.	48
Figura 22: Processo de tradução automática.	49
Figura 23: Detalhamento do módulo exibidor.	51
Figura 24: LibrasTV aplicada em sistemas de TV digital (componentes LibrasTV estão em destaque).	53
Figura 25: Visão esquemática do serviço proposto.	55
Figura 26: Arquitetura do sistema para geração automática de legendas em LIBRAS para cinemas digitais.	57
Figura 27: Plano de exibição e área de apresentação de LIBRAS.	65
Figura 28: Cena do filme Central do Brasil com <i>closed caption</i> , demonstrando a identificação do personagem e a emoção transmitida.	66
Figura 29: Cenas de filmes variados, com <i>closed caption</i> , demonstrando efeitos sonoros.	67

Figura 30: App LIBRAS-TV.....	70
Figura 31: Exemplo de imagens utilizadas na alfabetização de surdos, extraída de [44].	71
Figura 32: Imagens utilizadas para identificação do personagem.....	72
Figura 33: Identificação do personagem através da mudança de cor da roupa do intérprete.....	73
Figura 34: Demonstração do efeito sonoro (aplausos).	73
Figura 35: Demonstração de emoção sentida pelo personagem.....	74
Figura 36: Identificação do personagem através de imagens.	74
Figura 37: Largura de banda (em kpbs) utilizada pelo fluxo de LIBRAS (Video_1). ...	79
Figura 38: Largura de banda (em kpbs) utilizada pelo fluxo de LIBRAS (Video_2). ...	79
Figura 39: Largura de banda (em kpbs) utilizada pelo fluxo de vídeo com intérprete de LIBRAS (vídeo_3).	80
Figura 40: Gráfico sobre as diferenças regionais.	86
Figura 41: Gráfico sobre a identificação do personagem.	87
Figura 42: Gráfico que demonstra a opinião dos surdos sobre o grau de relevância do sistema proposto.	88
Figura 43: Gráfico que demonstra a aprovação dos surdos para o uso do sistema na ausência de um intérprete humano.	88
Figura 44: Grau de escolaridade dos usuários que participaram do teste.....	90
Figura 45: Fotos do experimento com os usuários.	91

Índice de Tabelas

Tabela 1: Valores possíveis para o campo <code>message_id</code> .	59
Tabela 2: Estrutura da <i>librasControlMessage</i> .	60
Tabela 3: Campo resolução.	61
Tabela 4: Estrutura da <i>librasDataMessage</i> .	63
Tabela 5: Possíveis valores para o <code>parameters_data_byte</code> .	67
Tabela 6: Estrutura do <i>Stream Event Descriptor</i> .	69
Tabela 7: Pistas visuais para sons do ambiente.	71
Tabela 8: Vídeos com <i>closed caption</i> utilizado no sistema.	78
Tabela 9: Resultados do teste de validação de hipótese para compreensão dos conteúdos	80
Tabela 10: Medida das médias de atraso de cada módulo do LibrasTV.	81
Tabela 11: Conteúdos multimídias usados na primeira bateria de testes.	82
Tabela 12: Conteúdos multimídias usados na segunda bateria de testes.	83
Tabela 13: Vídeos utilizados para avaliação das diferenças regionais.	85
Tabela 14: Vídeos utilizados para avaliar a identificação do personagem falante.	86
Tabela 15: Nível de conhecimento em português e em LIBRAS dos usuários.	90
Tabela 16: Níveis de compreensão para vídeos com <i>closed caption</i> .	91
Tabela 17: Níveis de compreensão para vídeos com janela de LIBRAS com avatar-3D.	92
Tabela 18: Resultados do teste de validação de hipótese para compreensão dos conteúdos multimídias.	93
Tabela 19: Auto-avaliação dos usuários para conhecimento em LIBRAS.	93
Tabela 20: Auto-avaliação dos usuários para conhecimento em língua portuguesa.	93
Tabela 21: Avaliação dos usuários para janela de LIBRAS com avatar-3D.	94
Tabela 22: Vídeos utilizados para identificação do personagem.	95
Tabela 23: Avaliação dos usuários sobre formas de identificação do personagem.	95

Lista de Acrônimos e Siglas

AAC – *Audio Advanced Encoding*
ACAP – *Advanced Common Application Plattform*
ATSC – *Advanced Television Systems*
ARIB – *Association of Radio Industries and Businesses*
BSL – *British Sign Language*
CPQD – *Centro de Pesquisas e desenvolvimento de Telecomunicações*
DSM-CC – *Digital Storage Media – Command and Control*
DVB – *Digital Video Broadcasting*
EPG – *Eletronic Program Guide*
ES – *Elementary Stream*
GEM – *Globally Executable MHP*
HDTV – *High Definition Television*
IBGE – *Instituto Brasileiro de Geografia e Estatística*
IHC – *Interação Humano Computador*
ISDB – *Integrated Services Digital Broadcasting*
ISDTV-T – *International Standard for Digital Television – Terrestrial*
ISO – *International Organization for Standardization*
LIBRAS – *Língua Brasileira de Sinais*
LS – *Língua de Sinais*
LSKB – *Língua Brasileira de Sinais Kaapor*
MHP – *Multimedia Home Platform*
MPEG – *Moving Picture Expets Group*
MPEG2-TS – *MPEG-2 Transport Stream*
NPT – *Normal Play Time*
OMS – *Organização Mundial de Saúde*
PAT – *Program Association Table*
PB – *Português Brasileiro*
PMT – *Program Map Table*
PSI – *Program Specific Information*
SBTVD – *Sistema Brasileiro de Televisão Digital*
UFPB – *Universidade Federal da Paraíba*

AGRADECIMENTOS

A Deus, por nunca me deixar desistir de lutar pelos meus sonhos, sempre me guiar nas decisões mais importantes da minha vida e ser minha fortaleza nos momentos mais difíceis.

Ao meu orientador Prof. Dr. Guido Lemos, por todo o apoio e ensinamentos.

Aos meus pais, pela preocupação com a minha educação desde garoto.

Ao meu irmão, Tony, pela compreensão e por suportar conviver comigo durante os períodos mais estressantes da pós-graduação.

À minha namorada, minha linda, meu amor, Polyane de Brito, por simplesmente existir.

À minha família.

A minha tia querida, Josilda, pela confiança, estímulo, mas, principalmente, por sempre me mostrar que sou capaz.

Em especial à minha mãe, Josélia Herminio Lemos, e ao meu pai, Ursulino Lemos Neto. Não existem palavras que possam descrever o que vocês representam para mim.

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. Além disso, 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, em que 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.

Palavras-chave: Televisão Digital, Língua Brasileira de Sinais, Conteúdos Multimídias, Protocolo, Diferenças Regionais.

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 over the video program. This solution is ineffective in several aspects: it has high operational costs, is dependent of a interpreter, and distract nondeaf viewers. In addition, it does not respects the regional differences in SLs. To reduce these problems, the main objective of this work 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.

Keywords: Digital Television, Brazilian Sign Language, Multimedia Contents, Protocol, Regional Differences.

1. INTRODUÇÃO

A comunicação é essencial na interação humana. Segundo Russell e Norvig [1], “a comunicação é a troca intencional de informações provocada pela produção e percepção de sinais extraídos de um sistema compartilhado de sinais convencionais”. Através desses sistemas compartilhados de sinais, denominados línguas (característica que nos difere dos demais seres), os humanos tem a oportunidade de expressar sentimentos, revelar conhecimentos, expor nossas opiniões frente aos assuntos relacionados ao nosso cotidiano, e, sobretudo, promovendo nossa inserção ao convívio social.

A língua utilizada por um indivíduo para se comunicar depende, em geral, do grupo no qual ele está inserido. Os ouvintes, por exemplo, se comunicam naturalmente através de línguas orais, isto é, através de sons articulados que são percebidos pelo sistema auditivo. As línguas inglesa, portuguesa, espanhola e francesa são exemplos de línguas orais. Os deficientes auditivos, por outro lado, se comunicam naturalmente através de línguas gestuais.

Os gestos são uma das principais ferramentas de suporte a comunicação. A família de movimentos que chamamos de gestos é bastante ampla, incluindo desde expressões faciais, passando por movimentos realizados com as mãos, até atividades realizadas com o corpo inteiro. Gestos associados à comunicação são denominados gestos semióticos, podendo ser classificados de acordo com a sua dependência com o discurso falado. Rimé e Schiaratura [2] propõem a seguinte subclassificação dos gestos semióticos quanto à sua funcionalidade:

- **simbólicos:** gestos cujo significado é único dentro de uma mesma cultura, como, por exemplo, o gesto de aprovação feito ao se exhibir a mão fechada apenas com o polegar voltado para cima. Linguagens de sinais também se enquadram nesta categoria;
- **deícticos:** gestos mais comumente utilizados em Interação Humano Computador (IHC), pois são aqueles utilizados para apontar ou direcionar a atenção a um determinado evento ou objeto;
- **icônicos:** estes são os gestos utilizados para transmitir informações quanto ao tamanho, forma ou orientação de um objeto em questão. Quando um pescador

diz: “Eu pesquei um bagre deste tamanho”, ao esticar seus braços lateralmente o máximo possível, ele está realizando um gesto semiótico icônico;

- **pantomímicos:** estes são os gestos realizados ao utilizarmos um instrumento ou objeto “invisível”, como, por exemplo, num jogo de mímica.

As duas modalidades de línguas, oral-auditiva (utilizada pelos ouvintes) e viso-espacial (linguagem gestual utilizada pelos surdos) são sistemas abstratos com regras gramaticais. Entretanto, da mesma forma que as línguas orais-auditivas não são iguais, variando de lugar para lugar, de comunidade para comunidade, as línguas de sinais também variam. Por exemplo, existe a língua de sinais americana, inglesa, francesa, brasileira e várias outras línguas de sinais em vários países.

A deficiência auditiva ou surdez consiste na perda parcial ou total da capacidade de ouvir, isto é, um indivíduo que possui algum problema auditivo. A surdez sendo de origem congênita, ou seja, quando o usuário nasce surdo, implica em uma série de dificuldades na aquisição da linguagem bem como no desenvolvimento da comunicação. A surdez adquirida é o resultado da perda de uma audição perfeita devido a lesões ou doenças. Nesta situação, na maior parte dos casos, a pessoa já aprendeu a se comunicar oralmente, porém, ao adquirir esta deficiência, terá que aprender a comunicar-se de outra forma.

Os deficientes auditivos possuem diversas particularidades, dentre elas, uma maior sensibilidade na compreensão e entendimento da comunicação motora, visual e espacial. Como utilizam língua diferente da adotada pelos ouvintes, ficam prejudicados quanto ao entendimento e compreensão da língua portuguesa brasileira tanto na sua forma oral quanto escrita. Uma vez que essas línguas possuem uma grafia baseada em sons, muitos deficientes auditivos passam vários anos na escola e não conseguem aprender a ler e escrever na língua oral de seu país [3]. Essa dificuldade resulta em uma grande barreira para a comunicação com outras pessoas, o acesso à informação, ao entretenimento, à aquisição de conhecimento, dentre outros.

Estudos empíricos fundamentados em acessibilidade mostram que indivíduos portadores de deficiências auditivas, seja ela parcial ou total, encontram muitos obstáculos para ter acesso definitivo aos conteúdos multimídia. Com o objetivo de minimizar os prejuízos encontrados na aquisição da informação e conhecimento pelos indivíduos surdos, o presente trabalho visa a elaboração de uma estratégia para

codificação das línguas de sinais com intuito de facilitar o acesso à comunicação, divertimento e aos conteúdos informativos, através de: cinemas digitais, com as legendas em língua de sinais apresentadas em dispositivos de *displays* externos, tais como *tablets*, *smartphones* e similares, que possuam recursos para o recebimento e a apresentação de fluxos de vídeo; como também, o acesso à informação e ao entretenimento disponível na televisão, (meio de comunicação com maior penetração nos domicílios brasileiros, 98%), com a exibição da janela com um intérprete de língua de sinais que pode ser tanto um humano quanto um agente virtual animado 3D.

1.1. Motivação

No Brasil existem duas línguas de sinais: a Língua Brasileira de Sinais (LIBRAS) e a língua Kaapor (LSKB). A LIBRAS é a língua de sinais utilizada pela maioria dos deficientes auditivos brasileiros e reconhecida pela lei brasileira nº 10436 de 24 de Abril de 2002 [4] como a língua de sinais oficial do Brasil. A língua de sinais Kaapor (LSKB), também conhecida como língua de sinais Urubú-Kaapor, é usada por uma pequena comunidade indígena da floresta amazônica brasileira. Os índios ouvintes dessa comunidade compreendem tanto a língua de sinais como a língua oral. Deste modo, os ouvintes da comunidade são encarados como bilíngues, enquanto os deficientes auditivos são monolíngues.

Segundo os dados apresentados pelo [IBGE] no censo realizado em 2010, o percentual de portadores de deficiências auditivas no Brasil é de 5,1%, classificadas em três categorias: (1) não consegue de modo algum – 0,18%, (2) grande dificuldade – 0,94% – e (3) alguma dificuldade – 3,98%. Aumento de 1,7% em relação ao censo realizado em 2000. 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 [5].

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. Uma

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 [6].

Além disso, para os surdos, as línguas orais, utilizadas nas legendas ocultas, representam apenas “uma segunda língua” [7]. 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 linguísticos manuais, corporais e faciais necessários para a articulação e significação visual-cultural do sinal [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 humano em tempo integral, (3) além de distrair os telespectadores que não são surdos e (4) não oferece suporte às diferenças regionais das línguas de sinais.

A língua oral portuguesa possui um enorme acervo no campo lexical e cada vez mais os léxicos disponíveis ao falante da língua portuguesa ampliam-se e estão em constantes modificações. Os modos diferentes de falar acontecem porque as línguas se transformam com o decorrer do tempo e sendo assim, ela assume características peculiares, em grupos sociais diferentes, em regiões diferentes, e com isso os indivíduos pertencentes a uma determinada comunidade aprendem aquela língua ou dialeto. As variações regionais, chamados dialetos, são as marcas determinantes referentes a diferentes regiões. Como exemplo, citamos a palavra mandioca que, em certos lugares, recebe outras nomenclaturas, tais como: **macaxeira e aipim**. As línguas de sinais, assim como as línguas orais, também podem diferir de uma região para outra em um mesmo país. Desta forma, o caráter de regionalização (sinal conhecido apenas em um determinado local que adota determinada língua de sinais) se torna muito importante e essencial para a fluência da comunicação. As Figuras 1 e 2 demonstram algumas formas de dialetos regionais presentes nas línguas orais e nas línguas de sinais.

Figura 1: Exemplos de variações regionais nas línguas orais.

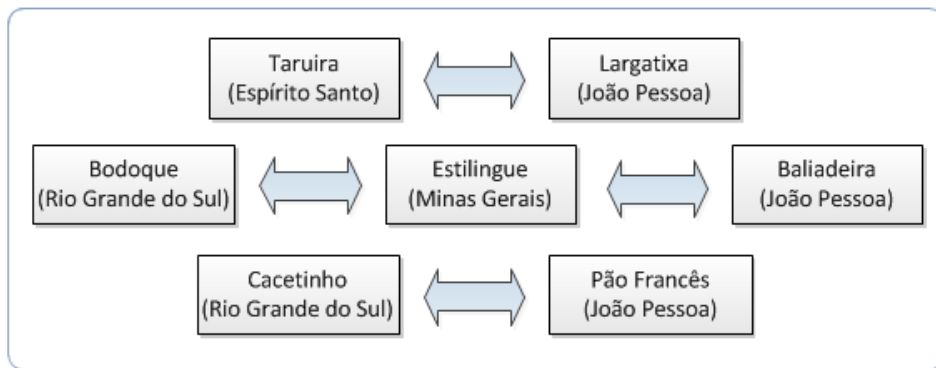


Figura 2: Exemplos de variações regionais nas línguas de sinais, extraída de [20].



Para minimizar esses problemas, definimos, como motivação desse trabalho, propor um protocolo que viabilize a codificação e transmissão da janela de LIBRAS em sistemas de TV Digital permitindo que as diferenças regionais sejam respeitadas. Adicionalmente incluiremos maneiras de representar efeitos sonoros, entonação de voz e nuances emocionais na sinalização, além de proporcionar uma redução na largura de banda utilizada para transmissão da janela de LIBRAS, pois, nessa proposta de codificação, os quadros do vídeo da janela de LIBRAS não são transmitidos. Em vez disso, um conjunto de códigos é transmitido, em que cada código está associado à

representação visual de um sinal de LIBRAS que está armazenada no receptor, conforme proposto por Araújo *et al.* [9] (ver Capítulo 5).

1.2. Objetivos

O objetivo geral deste trabalho é propor um protocolo de codificação da janela de LIBRAS para o Sistema Brasileiro de TV Digital (SBTVD). O protocolo possui suporte a representação de emoções e sons ambientes envolvidos em programas, pois, grande parte da semântica do cinema e da televisão é transmitida por meio das interações entre seres humanos na tela, emoções, som ambiente e música.

Para alcançar o objetivo geral definido por este trabalho, faz-se necessário que os seguintes objetivos específicos sejam alcançados:

- Fazer um levantamento bibliográfico detalhado sobre os principais trabalhos relacionados à codificação de língua de sinais e representação de emoções e efeitos sonoros para os surdos;
- Criar um protocolo para codificação da língua de sinais que possa contemplar os dialetos regionais;
- Definir a forma de representação das emoções e efeitos sonoros;
- Implementar um estudo de caso para a validação do protocolo;
- Planejar e definir os experimentos com os deficientes auditivos;
- Testar e validar essa proposta com deficientes auditivos e intérpretes de LIBRAS.

1.3. Escopo do Trabalho

O problema abordado neste trabalho possui natureza fortemente interdisciplinar, envolvendo as áreas de Acessibilidade, Computação Gráfica, Inteligência Artificial, Sistemas Multimídia, dentre outras. Dessa forma, o tratamento pleno de todos os aspectos do problema não é trivial. Portanto, restringimos o problema a um escopo tratável em que foram considerados os seguintes aspectos:

1. Apenas mídias textuais serão usadas como base para geração dos conteúdos acessíveis. Os conteúdos acessíveis serão gerados apenas a partir de mídias baseadas em texto como legendas, *closed caption*, entre outros.
2. A proposta do trabalho não é substituir ou se equiparar aos intérpretes humanos. Atualmente, a qualidade da sinalização virtual ainda não se aproxima da qualidade da sinalização humana. Dentre as razões para esse distanciamento pode-se destacar a dificuldade das estratégias de sinalização virtual¹ de representar emoções e movimentos de forma tão natural quanto os humanos. Em consequência disso, a proposta não é substituir intérpretes humanos ou gerar traduções tão boas quanto às geradas por intérpretes humanos. A proposta é desenvolver uma solução complementar, prática e de baixo custo que possa ser utilizada para reduzir as barreiras de comunicação e acesso a informação dos surdos em conteúdos televisivos, especialmente quando intérpretes humanos não estão disponíveis ou não são viáveis.

1.4. Organização do Trabalho

Esse trabalho está estruturado em sete capítulos. Este primeiro capítulo apresenta uma motivação inicial sobre o trabalho, apresentando conceitos relevantes como, línguas de sinais e sua utilização e importância na comunicação dos deficientes auditivos.

No segundo capítulo são apresentados alguns conceitos básicos sobre TV digital e o SBTVD. Também serão apresentadas informações mais detalhadas sobre a Língua Brasileira de Sinais (LIBRAS), foco do trabalho em questão.

No terceiro capítulo são apresentados os principais trabalhos da literatura científica relacionados ao tema do trabalho proposto.

¹ Neste trabalho, o termo sinalização virtual é usado para representar as estratégias que utilizam agentes animados virtuais (isto é, avatares) para representação dos sinais em línguas de sinais.

No quarto capítulo são apresentados os trabalhos relacionados à acessibilidade desenvolvidos na Universidade Federal da Paraíba (UFPB), especificamente no Laboratório de Aplicações de Vídeo Digital (LAVID).

No quinto capítulo é apresentada uma descrição detalhada sobre o protocolo proposto nesse trabalho.

O sexto capítulo apresenta o plano de experimentos e os testes utilizados para validar a solução proposta, como também experimentos realizados com usuários surdos e intérpretes de LIBRAS.

No sétimo capítulo são expostas as conclusões e considerações finais sobre o trabalho, além de algumas sugestões de atividades a serem desenvolvidas no escopo desse trabalho.

A lista de referência fecha o trabalho, com o material que serviu de suporte para o mesmo.

O anexo A apresenta os questionários utilizados para avaliação dos surdos.

O anexo B apresenta as produções técnicas e científicas envolvendo o protocolo proposto, no escopo do projeto LibrasTV.

2. LIBRAS NO SISTEMA BRASILEIRO DE TV DIGITAL

Nesse capítulo são apresentados os principais conceitos necessários para melhor compreensão desse trabalho. Inicialmente são apresentados alguns conceitos relacionados aos Sistemas de TV Digital, que incluem algumas tecnologias, protocolos e estatísticas relevantes para a contextualização do trabalho. Em seguida são apresentadas algumas informações básicas sobre a Língua Brasileira de Sinais (LIBRAS).

2.1. TV Digital

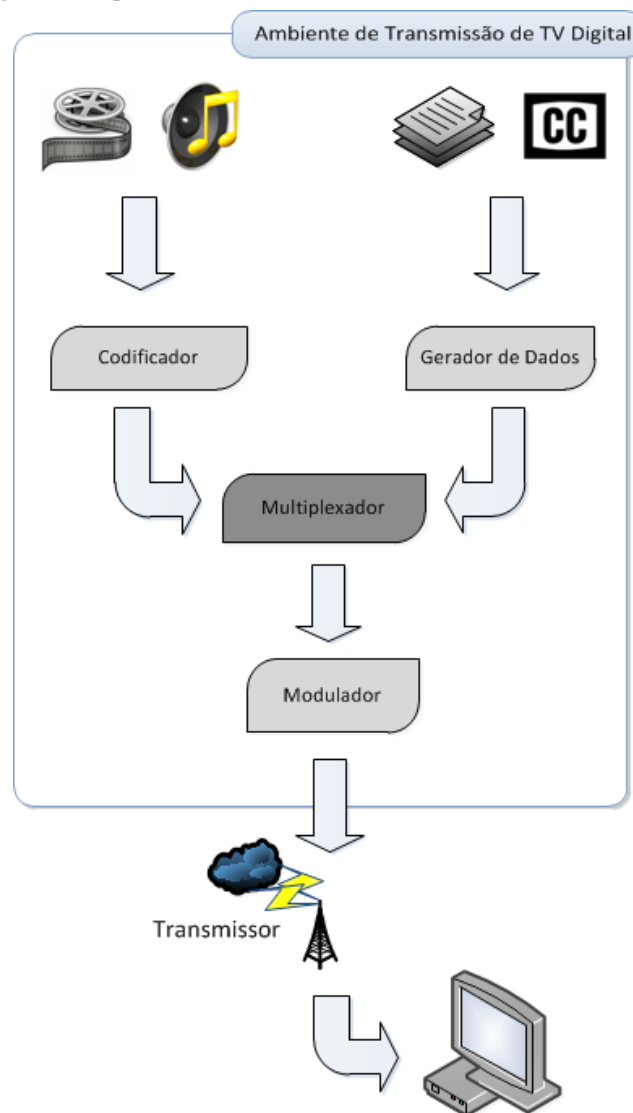
Na TV Digital, os sinais de áudio e vídeo são digitalizados e codificados utilizando algoritmos específicos para compressão, o que permite um uso mais eficiente do espectro eletromagnético, permitindo dessa forma:

- Melhoria de qualidade de som e imagem, viabilizando um sinal de alta definição (ou Televisão de Alta Definição - HDTV) com resolução de até 1920 x 1080 pixels;
- Possibilidade de transmitir simultaneamente mais de um programa dentro do mesmo canal físico (multiprogramação);
- Além disso, tecnologias de transporte de dados permitem que fluxos de dados sejam multiplexados com os sinais de áudio e vídeo. Como exemplos de fluxos de dados, podemos ter:
 - Informações relacionadas aos programas de TV, como, por exemplo, informações relacionadas do radiodifusor, frequência do canal, etc;
 - Guia Eletrônico de Programação (*Electronic Program Guide* – EPG);
 - Aplicações interativas que permitem que o usuário interaja com a emissora através de um canal de retorno. Dessa forma, o usuário pode votar ou fazer compras através da TV;
 - Legendas Ocultas (*Closed caption* – CC);

2.1.1. Sistemas de TV digital

Um ambiente de transmissão de TV Digital é composto pela Estação de TV, responsável por transmitir o sinal para os telespectadores, e um receptor desses sinais, (aparelho de TV, *set-top-box*, computador, celular, etc), responsável por receber o sinal transmitido, decodificá-lo e apresentá-lo ao usuário. A Figura 3 apresenta os elementos de um ambiente de transmissão de TV Digital.

Figura 3: Arquitetura de um ambiente de transmissão de TV Digital.



Inicialmente, os sinais de vídeo e o áudio são capturados por uma câmera, digitalizados e comprimidos utilizando um Codificador (*Encoder*). Paralelamente,

fluxos de dados, como informações adicionais sobre o programa, EPGs, aplicações interativas e *closed caption* são gerados e codificados pelo Gerador de Dados. Em seguida, os fluxos codificados de áudio, vídeo e dados são multiplexados (pelo Multiplexador) em um fluxo de transporte único, em geral, compatível com o padrão MPEG-2 Sistemas. O fluxo de transporte é então modulado (pelo Modulador) e transmitido numa rede de difusão.

Do outro lado, no receptor, o sinal da TV é sintonizado, demodulado, decodificado e apresentado ao usuário. O receptor também possui a capacidade de executar as aplicações interativas recebidas conjuntamente com o vídeo. Esse receptor pode ser visto, portanto, como um computador adaptado para as necessidades do ambiente de TV, possuindo processador, memória, sistema operacional, etc. Ele também executa um *middleware*², cuja função é abstrair as características específicas de cada receptor, permitindo que a mesma aplicação seja executada em receptores construídos por diferentes fabricantes [10].

A função de cada elemento do ambiente de transmissão de TV (ver Figura 3) pode ser resumida da seguinte forma:

- **Codificador:** Converte os sinais analógicos de áudio e vídeo para o formato digital de acordo com padrões de compressão, como o MPEG-2, H.264, MPEG-2 AAC (*Audio Advanced Encoding*);
- **Gerador de Dados:** Elemento responsável por gerar e codificar fluxos de dados como informações adicionais sobre o programa, EPG, aplicações interativas, *closed caption*, etc;
- **Multiplexador:** Responsável por fundir os fluxos de áudio, vídeo e dados em um único fluxo de transporte. O mecanismo de transporte de dados adotado pela maioria dos Sistemas de TV Digital é o MPEG-2 Sistemas;
- **Modulador:** Modula o fluxo de transporte oriundo do multiplexador utilizando um determinado padrão de modulação;
- **Receptor:** Decodificam os sinais de áudio, vídeo e *closed caption*, gerencia e executa as aplicações interativas. Pode ser definido, como um computador adaptado para as necessidades do ambiente de TV, controlado por um

² O middleware é uma camada de software responsável por abstrair as características específicas de cada receptor, permitindo que o mesmo aplicativo interativo possa ser executado em receptores de diferentes fabricantes [10]

middleware que atua em uma camada de abstração acima do sistema operacional.

Para tornar compatíveis os elementos de transmissão e recepção dentro de um ambiente de transmissão de TV Digital, é necessário especificações que normatizam todo o processo de captura, compressão, modulação e transmissão dos sinais de vídeo, bem como todas as interfaces físicas entre os equipamentos envolvidos no processo, de forma a serem transmitidos. Um fluxo de transporte é formado por um ou mais fluxos elementares (ES – *Elementary Stream*). Um *fluxo elementar* é definido como um fluxo de dados gerado pela codificação do conteúdo de vídeo, áudio ou outros dados.

As especificações MPEG-2 Sistemas definem ainda o termo *programa*, chamado de *serviço* no contexto da TV digital, como um grupo composto de um ou mais fluxos elementares com uma mesma base temporal [11]. O fluxo de transporte pode conter vários serviços (programas) simultaneamente, cada um podendo ter uma base de tempo diferente.

Simplificadamente, multiplexar serviços em um fluxo de transporte significa organizar os pacotes dos vários fluxos elementares, pertencentes aos serviços contemplados, em um único fluxo. Para isso, é necessário inserir no fluxo de transporte, informações que permitam ao decodificador MPEG-2 identificar a qual serviço um dado fluxo elementar pertence. Essas informações são dispostas como um conjunto de tabelas de informação específica de programa (PSI – *Program Specific Information*). Uma PSI particular, denominada PMT (*Program Map Table*), contém a lista de identificadores dos fluxos elementares que compõem um serviço. Cada PMT encontrada representa um serviço disponível. As PMTs são localizadas através de outra tabela do PSI denominada PAT (*Program Association Table*), que contém identificadores dos fluxos elementares contendo as PMTs. O fluxo elementar que possui a PAT possui identificador fixo com o valor hexadecimal 0x00.

2.2. Eventos de Sincronismo DSM-CC

Sincronizar o comportamento de uma aplicação com o conteúdo de uma programação de TV específica (áudio e vídeo principal) é extremamente desejável, principalmente quando a aplicação em questão possui relação semântica com essa

programação. O sincronismo pode ser realizado através de *selos de tempos* (paradigma *timeline*), quando os dados das aplicações são enviados de forma síncrona, ou sincronizada [11] com o áudio e vídeo principal, ou alternativamente através de *eventos de sincronismo* DSM-CC (ou simplesmente eventos DSM-CC), quando os dados são transmitidos de forma assíncrona (sem o uso de selos de tempo, utilizando, por exemplo, o carrossel de objetos).

Para criar um evento DSM-CC, é inserida uma estrutura no fluxo de transporte, denominada descritor de eventos. Cada *descritor de eventos* possui um identificador numérico único, que o identifica no fluxo de transporte. Uma vez que não existe a possibilidade do provedor de conteúdo saber exatamente a posição que o descritor será inserido no fluxo, cada descritor possui uma referência temporal que indica em qual instante o evento deverá ocorrer, usualmente baseado em um fluxo DSM-CC denominado NPT (*Normal Play Time*) [12]. Como caso particular, um descritor de eventos pode informar ao sistema receptor que o evento deve ocorrer imediatamente, esse tipo de evento é chamado de evento “*do it now*”. Além do identificador e da referência temporal, o descritor de eventos possui também um campo para dados específicos das aplicações, que pode ser utilizado de acordo com sintaxes e semânticas a serem tratadas pelas próprias aplicações. Para evitar que um descritor de evento DSM-CC seja perdido devido a falhas no meio de difusão, um mesmo descritor normalmente é enviado diversas vezes pelo provedor de conteúdo, cabendo ao receptor interpretar esses eventos uma única vez.

Os descritores de eventos DSM-CC, bem como os outros dados DSM-CC (por exemplo, dados referenciados pelos eventos e que devem ser sincronizados com o áudio e vídeo principal), precisam de estruturas especiais para ser transportados. Para tanto, as especificações do protocolo DSM-CC determinam estruturas de dados denominadas *seções* DSM-CC. Essas estruturas possuem um cabeçalho, especificado com o objetivo de informar ao decodificador como as seções estão sendo utilizadas para transportar os dados, como elas devem ser remontadas, quais os tipos de dados transportados, além de outros parâmetros para o tratamento apropriado da informação [12].

2.3. Sistema Brasileiro de TV Digital (SBTVD)

No ano de 2005 o governo brasileiro instituiu o projeto SBTVD (Sistema Brasileiro de TV Digital), do qual participaram milhares de pesquisadores de várias universidades e centros de pesquisa. As atividades do SBTVD abordaram todas as partes que compõem um sistema de TV Digital, além do desenvolvimento de inovações que objetivaram ajustar a implantação da TV Digital ao contexto social brasileiro.

O objetivo do projeto foi oferecer ao governo as informações necessárias para guiá-lo no sentido da criação de um sistema aberto de televisão digital terrestre, posteriormente batizado como ISDTV-T (*International Standard for Digital Television – Terrestrial*).

O *middleware* de referência, denominado Ginga, teve sua origem na junção do *middleware* procedural FlexTV [14], desenvolvido em compatibilidade com o GEM [15], e o *middleware* declarativo MAESTRO [16], ambos concebidos no projeto em questão.

Os requisitos que conduziram a especificação do Ginga foram, em sua maioria, baseados em algumas particularidades do contexto social do país, as quais são compartilhadas pela maioria dos países latino-americanos. Assim, algumas inovações em relação aos outros *middlewares* se fizeram necessárias e foram implementadas preservando a compatibilidade com estes.

No mesmo ano (2005) da instituição do SBTVD, o CPqD (Centro de Pesquisas e desenvolvimento de Telecomunicações) verificou [17] que havia por volta de 55 milhões de televisores presentes em 90% dos lares brasileiros; dessa porcentagem, 79% só recebiam via radiodifusão terrestre. Quanto aos equipamentos utilizados, 27% eram monitores de 14 polegadas, 37% de 20 polegadas, com 47% deles exclusivamente com antena interna.

Estatísticas como essas, somadas ao fato de apenas 32,1 milhões de brasileiros terem acesso à Internet (21% da população), fizeram o governo determinar alguns requisitos básicos para o Sistema Brasileiro de TV Digital. Entre eles tem-se: baixo custo e robustez na recepção, flexibilidade e capacidade de evolução, interatividade e novos serviços. Este visa promover a inclusão digital e é tratado como um requisito fundamental.

2.4. Língua Brasileira de Sinais

Conforme mencionado anteriormente, as línguas de sinais são a forma natural de comunicação utilizada pelos deficientes auditivos. Segundo Brito [18], elas são consideradas línguas naturais, pois surgem espontaneamente da interação entre os deficientes auditivos e podem expressar qualquer conceito descritivo, concreto, racional, literal, metafórico, emocional ou abstrato.

A LIBRAS, assim como toda língua natural, possui uma estrutura gramatical bem definida. Ela é construída a partir de parâmetros de configuração (fonemas), que são estruturados a partir de mecanismos morfológicos, sintáticos e semânticos. Os princípios pragmáticos, presentes também na LIBRAS, permitem a criação de alguns tipos de discursos: metafóricos, ironias, dentre outros.

Brito [18] e Quadros e Karnopp [19] também ressaltam que existe uma diferença entre as línguas orais e as línguas de sinais com relação à estrutura de apresentação ao longo do tempo. Enquanto nas línguas orais, os fonemas se sucedem linearmente no tempo, possuindo uma estrutura sequencial, nas línguas de sinais, cada sinal pode utilizar diversas regiões do corpo do sinalizador simultaneamente, levando a uma estrutura paralela de representação dos sinais.

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 [20]:

- **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, a Figura 4 demonstra um exemplo de configuração da mão. Existem atualmente 60 possíveis configurações de mão, como podemos ver na Figura 5;

Figura 4: Exemplo de configuração de mão em sinais, extraída de [21].

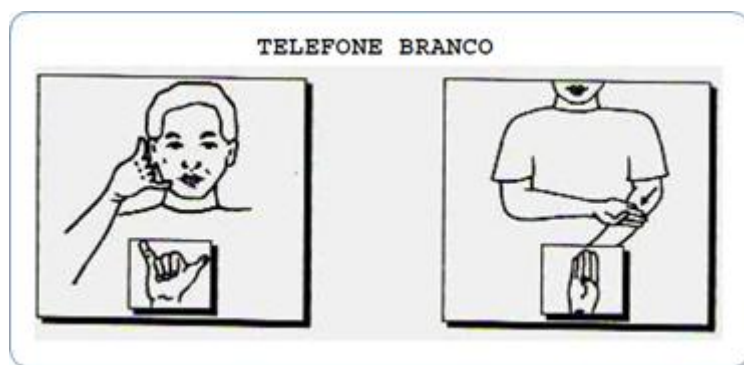
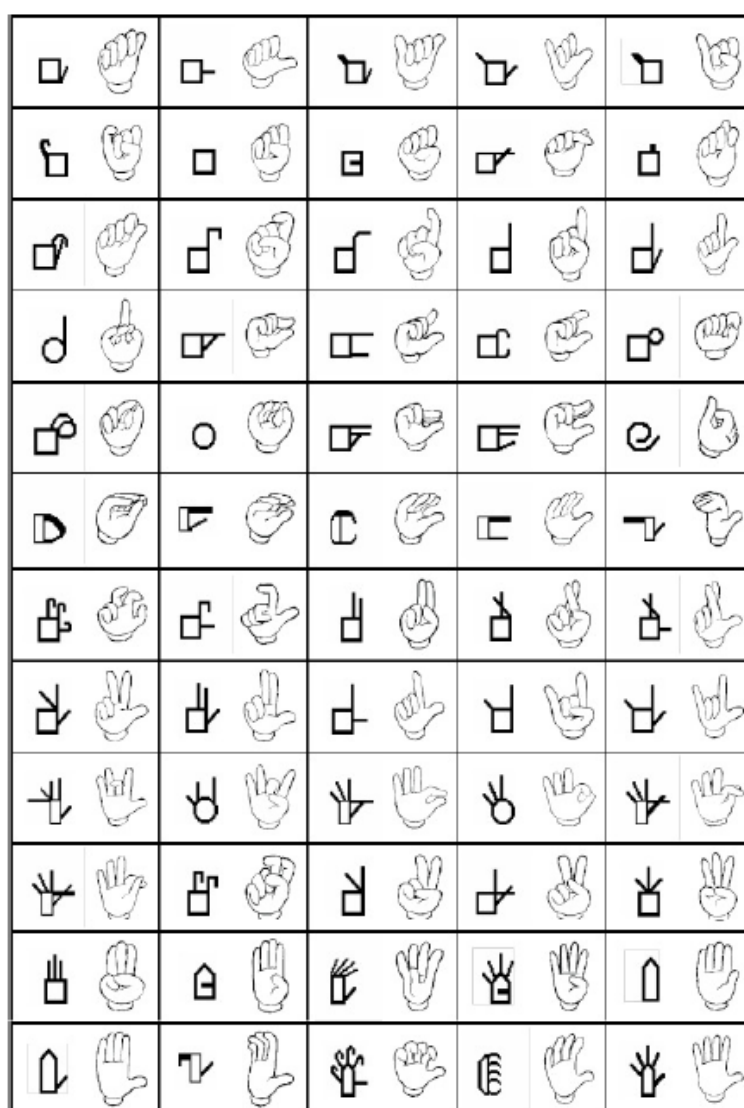


Figura 5: Configurações de mão da LIBRAS, extraída de [21].



- **ponto de articulação:** representa a parte do corpo do emissor onde os sinais são realizados (ou começam a ser realizados). Em LIBRAS, esse ponto de articulação pode estar localizado em alguma parte do corpo ou estar

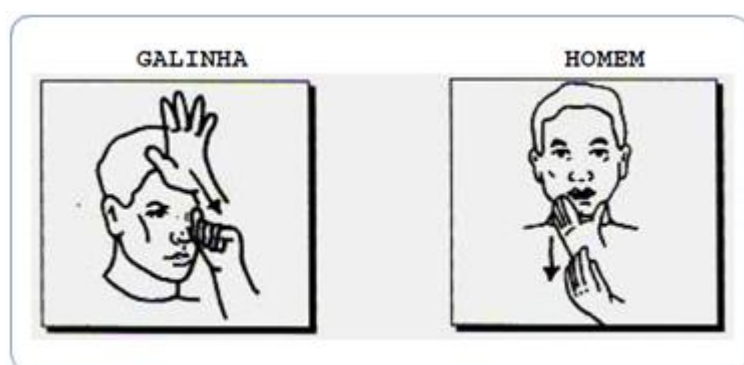
localizado em um espaço neutro vertical (do meio do corpo até a cabeça) e horizontal (à frente do emissor), Figura 6;

Figura 6: Exemplo de pontos de articulação, extraída de [21].



- **movimento:** um sinal pode ter movimentos ou não. O movimento representa o deslocamento da mão no espaço durante a execução do sinal. Caracterizar os movimentos é complicado, uma vez que a mão é um objeto assimétrico. Além disso, os eixos podem se deslocar simultaneamente. Segundo Fusco [21], no entanto, a maioria dos sinais pode ser dividida em pequenos segmentos de movimentos e cada um dos quais relacionados a um eixo. Em LIBRAS, por exemplo, os sinais AJOELHAR e EM-PÉ não tem movimento, podemos observar um exemplo de movimento na Figura 7;

Figura 7: Exemplo de movimento, extraída de [21].



- **expressão facial e/ou corporal:** muitos sinais podem requerer características adicionais para expressar sentimentos de alegria, de tristeza, uma pergunta ou uma exclamação e necessitam de um traço diferenciador, como uma expressão facial e/ou corporal, para poder expressar essas

características. Dessa forma, as expressões faciais e/ou corporais podem assumir tanto uma função léxica (expressar alegria e tristeza, por exemplo) quanto uma função sintática (expressar uma pergunta ou uma exclamação, por exemplo) na estrutura dos sinais. A inclusão de uma expressão facial e/ou corporal no discurso pode alterar o significado de um sinal e sua ausência pode produzir um sinal sem sentido. Alguns sinais em LIBRAS, inclusive, são feitos apenas com a bochecha, como LADRÃO e ATO-SEXUAL, na Figura 8 podemos ver dois exemplos de expressões faciais;

Figura 8: Exemplo de expressões faciais, extraída de [21].



- **orientação/direção:** Os sinais têm uma direção com relação aos parâmetros acima. Assim, os verbos IR e VIR se opõem em relação à direcionalidade, a Figura 9 dá um exemplo da orientação/direção dos sinais.

Figura 9: Exemplo da orientação dos sinais, extraída de [21].



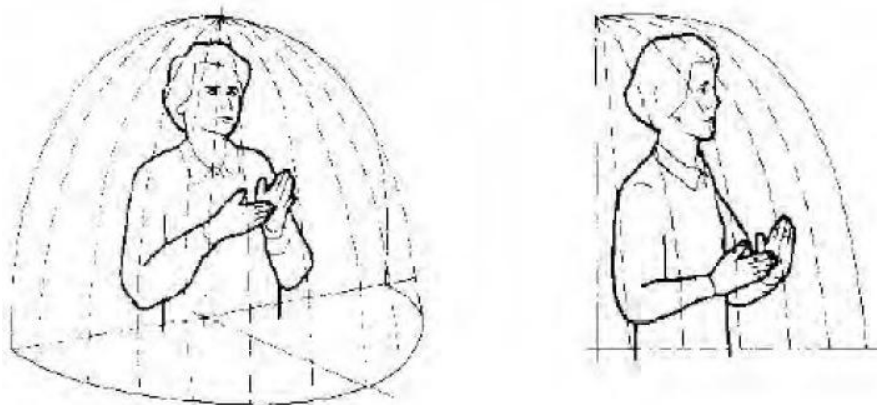
A partir da combinação desses cinco parâmetros formam-se os sinais, que permitem comunicar diferentes ideias ou aspectos de uma sentença. Expressar-se em

LIBRAS ou em qualquer outra língua de sinais, portanto, é combinar esses elementos para formar os sinais e combinar os sinais para formar sentenças em um determinado contexto.

2.4.1. Espaço de sinalização

O espaço de sinalização corresponde à área em que os sinais são articulados nas línguas de sinais. Essa área é um espaço tridimensional (3D) que vai da cabeça até a cintura e se estende para fora do corpo até que o braço esteja completamente esticado. As articulações de mãos e expressões faciais ou corporais de um sinal ocorrem sempre nesse espaço. Um diagrama desse espaço, extraído de [22], é apresentado na Figura 10.

Figura 10: Diagrama do espaço de sinalização.



Na comunicação em línguas de sinais, esse espaço de sinalização pode ser dividido nos pontos de referência dêitico e no espaço de sinalização neutro. Os pontos de referência dêíticos podem ser usados por um sinalizador (isto é, uma pessoa que está representando os sinais) para associar pessoas, objetos ou conceitos em discussão com locais (ou pontos de referência) no espaço de sinalização. Por exemplo, suponha uma cena em que exista uma casa no lado esquerdo. Nesse caso, um sinalizador pode representar o sinal “casa” e apontar para o lado esquerdo no espaço de sinalização. Isso indica que um ponto de referência para o sinal “casa” foi criado no espaço de sinalização. Se o sinalizador quiser representar um homem andando em direção a casa, ele pode simplesmente utilizar esse ponto de referência para indicar que a casa é o ponto final do movimento, isto é, ele não precisa representar o sinal “casa” novamente e pode

utilizar seu ponto de referência. A utilização desses pontos de referência no espaço 3D não ocorre nas línguas orais e representa uma importante informação semântica para as sentenças.

O espaço neutro, por outro lado, representa todos os outros locais no espaço de sinalização que não contém pontos de referência associados.

2.5. Considerações Finais

A partir das informações contidas neste capítulo, pode-se ter uma melhor compreensão sobre o desenvolvimento desse trabalho. Foram apresentados alguns conceitos relacionados aos Sistemas de TV Digital como também ao (SBTVD), que incluem algumas tecnologias relevantes para a contextualização do trabalho. Por fim foi descrito os principais conceitos para o entendimento língua de sinais foco desse trabalho, a Língua Brasileira de Sinais (LIBRAS). No próximo capítulo são apresentados os principais trabalhos presentes na literatura científica, relacionados ao tema proposto.

3. TRABALHOS RELACIONADOS

Nesse capítulo são apresentadas revisões sobre os principais trabalhos da literatura relacionados com a transmissão da língua de sinais como também a representação de efeitos sonoros, entonação de voz e nuances emocionais para os deficientes auditivos. Posteriormente, conforme proposto nesse trabalho, é apresentado um protocolo que viabiliza 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.

3.1. Representação de Emoções e Sons Ambientes para Deficientes Auditivos

Nesta seção serão expostos alguns trabalhos presentes na literatura científica relacionados à representação de emoções e efeitos sonoros para deficientes auditivos através do *closed caption*.

3.1.1. *Emotive Captioning*

O *Emotive Captioning*, proposto por Rashid *et al.* [23] busca a formulação de legendas emotivas, constituídas por três características importantes: (1) que combine elementos gráficos e texto para representar efeitos de emoção e som; (2) trabalhar em colaboração com uma equipe de produção para identificar as emoções existentes em um programa de televisão, e (3) trabalhar com uma equipe de *design* para criar e implantar as legendas gráficas.

O trabalho de Rashid *et al.* propõe a utilização de 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 utilizadas para representar emoção (ver Figura 11).

Figura 11: Legenda emotiva dinâmica, para representar engarrafamento no trânsito, extraída de [23].



Uma das dificuldades encontradas por Rashid em utilizar imagens para transmitir emoções é decidir qual a informação deve ser transmitida, a melhor forma de transmití-la e a imagem mais adequada para exibição. Estas dificuldades possuem um potencial grande de se tornarem um problema complicado. Com isso, legendadores seriam obrigados a vasculhar grandes bases de dados ou criar suas próprias imagens para cada novo projeto de legendas. Este processo iria assegurar ao escritor responsável pela legenda, a interpretação ao invés de uma tradução literal. Pensando nesse problema, para simplificar o vocabulário emotivo, seis emoções foram colocadas à disposição do legendador: medo, raiva, tristeza, felicidade, repugnância e surpresa (como citado acima). Também foram adicionados duas emoções, "luxúria" e "amor", juntamente com uma escala de 1 a 5 para medir o nível de intensidade para cada emoção. Uma das limitações dessa solução é a falta de avaliação por usuários.

3.1.2. *Emotional Subtitles*

O *Emotional Subtitles*, proposto por Ohene-Djan *et al.* [24], 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 12. 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.

Figura 12: O programa de televisão BBC's Tweenies com legenda emocional, extraída de [24].



3.1.3. *Dynamic Captioning*

O *Dynamic Captioning*, proposto por Richang Hong *et al.* [25], investiga uma abordagem automática para apresentar de forma inteligente as legendas. A Figura 13 ilustra os diferentes estilos de legendas.

Figura 13: Exemplos de diferentes estilos de legendas: a) legendas scroll-up, b) legendas pop-up, c) legendas pain-on, d) legendas cinematográficas, e) legendas dinâmicas, extraído de [25].



Conforme pode ser observado na Figura 13, as quatro primeiras técnicas (Figura 13a, 13b, 13c e 13d) podem ser categorizadas como legendas estáticas, diferentemente da última técnica (Figura 13e) classificada como legenda dinâmica. Segundo Richang Hong *et al.*, os deficientes auditivos são beneficiados com a utilização desse tipo de legenda, uma vez que as legendas são apresentadas de forma síncrona com a fala do personagem realçando palavra por palavra e ilustrando a variação de tonalidade da voz do personagem. Na Figura 14 podemos observar a Figura 13e ampliada, para uma melhor visualização do posicionamento das legendas dinâmicas.

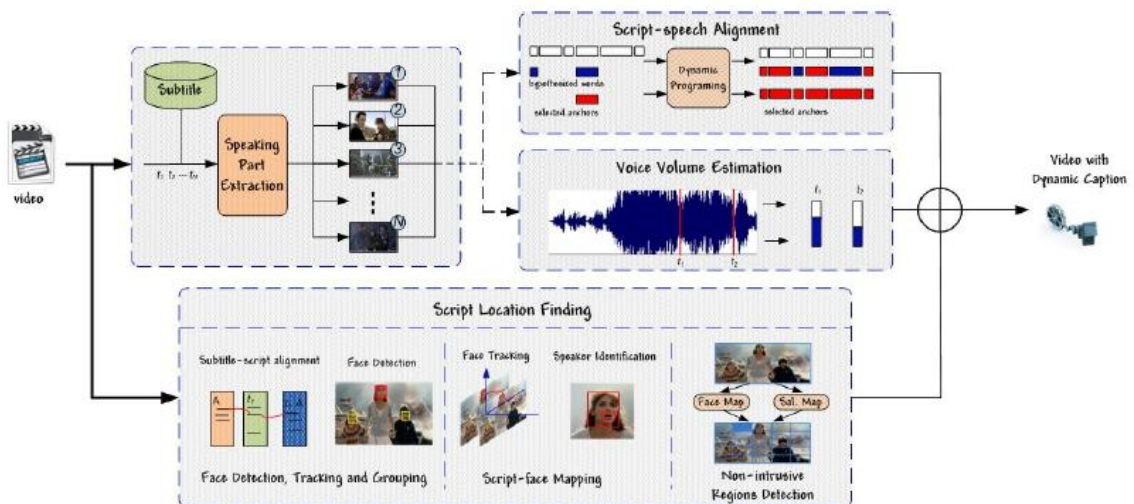
Para o processo de criação do vídeo acessível, segundo Richang Hong *et al.*, inicialmente são extraídos partes das falas dos personagens de acordo com as informações de tempo das legendas. Em seguida os personagens são mapeados com as legendas correspondentes com o auxílio de técnicas de reconhecimento e detecção de face. Posteriormente, uma região não-intrusiva é detectada ao redor do rosto do personagem com base na análise da saliência visual em que as legendas são apresentadas. Em paralelo, as legendas estão alinhadas com a faixa de áudio baseado no roteiro de fala do personagem. Com base nessas informações, as legendas são destacadas sincronicamente, juntamente com o discurso do personagem, palavra por palavra, para que o público deficiente auditivo possa melhor visualizá-las. As legendas estão localizadas perto dos personagens para ajudar a platéia surda a identificar a emoção dos personagens correspondentes.

Figura 14: Exemplo de legenda dinâmica, extraída de [25].



Uma das limitações desse trabalho é a imprecisão no mapeamento da face dos interlocutores da cena. A Figura 15 apresenta uma ilustração esquemática do trabalho de Richang Hong *et al.*

Figura 15: Ilustração esquemática da proposta de Richang Hong et al., extraído de [25].



3.2. Transmissão de Línguas de Sinais

Nesta seção são apresentados alguns trabalhos presentes na literatura científica relacionada à transmissão da língua de sinais.

3.2.1. *MobileASL*

O MobileASL [26] é um projeto de compressão de vídeo da Universidade de Washington e da Universidade de Cornell [27], e tem como objetivo principal, tornar a comunicação por telefone celular para pessoas que usam línguas de sinais uma realidade nos EUA. O software desenvolvido por Jaehong Chon *et al.* [28] é executado no sistema operacional Windows Mobile. O MobileASL é compatível com o padrão de compressão H.264/AVC e pode ser decodificado por qualquer decodificador H.264. No sistema proposto por Jaehong Chon *et al.* foi utilizado o protocolo UDP para transmissão de vídeo e o HTTP para o registro de informações do usuário (como o endereço IP) ao servidor, que está localizado fora da rede.

No projeto MobileASL foi implementado um sistema de codificação da região de interesse, baseado na detecção da pele, para reduzir a largura de banda necessária, mantendo a qualidade e inteligibilidade necessária para o entendimento da ASL. O sistema proposto por Jaehong Chon *et al.* é capaz de comprimir, transmitir e decodificar 12-15 frames por segundo em tempo real e produzir ASL inteligível em 30 kbps. Uma das limitações desse sistema é a possibilidade de ocorrência de atrasos significativos que podem prejudicar o entendimento da língua de sinais (LS) como também a falta de suporte ao regionalismo. A Figura 16 demonstra um *screenshot* do MobileASL em funcionamento.

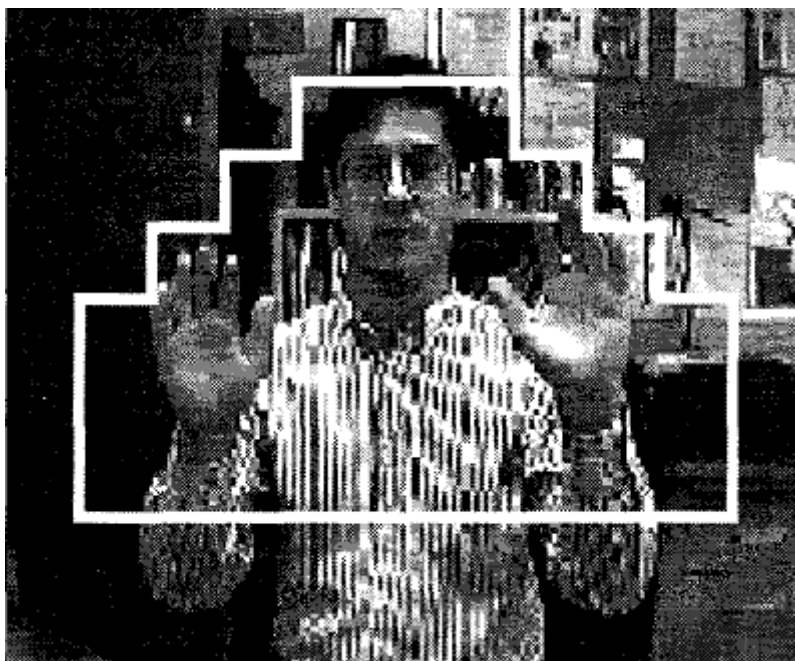
Figura 16: Screenshot do codec MobileASL sobre o celular HTC TyTN II, extraído de [28].



3.2.2. Prioridade na codificação da região de interesse para língua de sinais

O trabalho proposto por Richard P. Schumeyer *et al.* [29] investiga a compressão de sequências de vídeo com língua de sinais para que sejam transmitidos em canais com baixas taxas de bits. O movimento das mãos e braços são inerentes à linguagem sinalizada e requerem alta resolução temporal e uma maior compressão para os canais de banda limitada. A maioria dos esquemas de compressão existente trata toda a imagem de maneira uniforme, podendo comprometer o entendimento de expressões faciais e movimentos dos dedos. O algoritmo proposto nesse trabalho dá prioridade ao primeiro plano, sacrificando a qualidade do fundo para alcançar taxas de compressão mais altas sem comprometer a taxa de quadros. O trabalho não oferece nenhum suporte ao regionalismo, porém fornece uma alternativa para diminuir o espaço utilizado na banda. A Figura 17 demonstra a região de interesse que será extraída de um vídeo para posterior compressão e codificação.

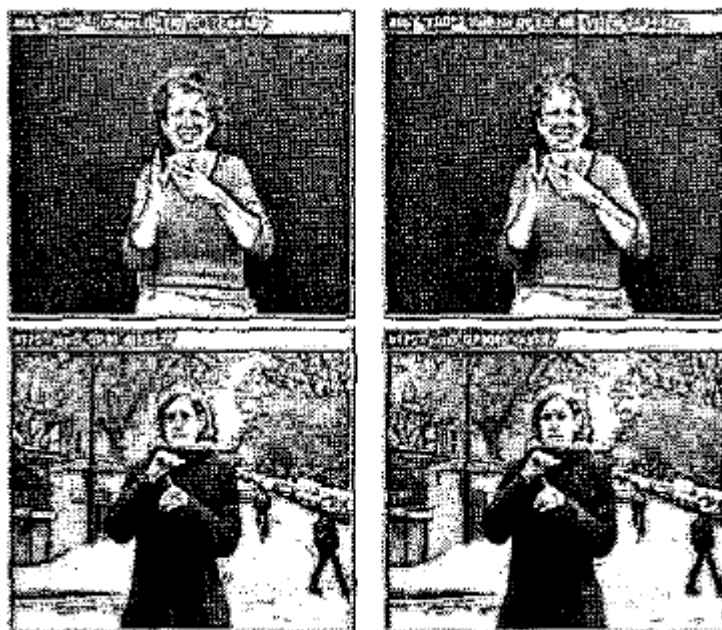
Figura 17: Região de interesse para codificação, extraído de [29].



3.2.3. Sistema de codificação de vídeo para comunicação da língua de sinais a baixas taxas de bits

O sistema proposto por Dimitris Agrafiotis *et al.* [30], similar ao trabalho de Richard P. Schumeyer *et al.*, propõe uma maneira de transmitir a língua de sinais em baixas taxas de bits. A abordagem de codificação utilizada nesse trabalho visa remover componentes de alta frequência espacial para reduzir a entropia do vídeo permitindo maior ganho na compressão. Segundo Dimitris Agrafiotis *et al.* devem ser codificadas apenas as regiões de interesse dos deficientes visuais (rosto e braços). Partes estáticas do vídeo serão codificadas apenas uma vez. A fim de avaliar o efeito da abordagem proposta, sobre a capacidade de compreensão do vídeo codificado, foi feito um teste com 17 surdos. Os participantes mais experientes aprovaram o sistema, porém, os demais, que não possuíam fluência com a BSL, não compreenderam o que estava sendo transmitido, pois, não houve preservação da qualidade do vídeo. A Figura 18 demonstra os resultados da codificação com o método proposto por Dimitris Agrafiotis *et al.*, o plano de fundo é codificado apenas uma vez.

Figura 18: Resultados da codificação com o método proposto, extraído de [30].



3.2.4. *Rybená*

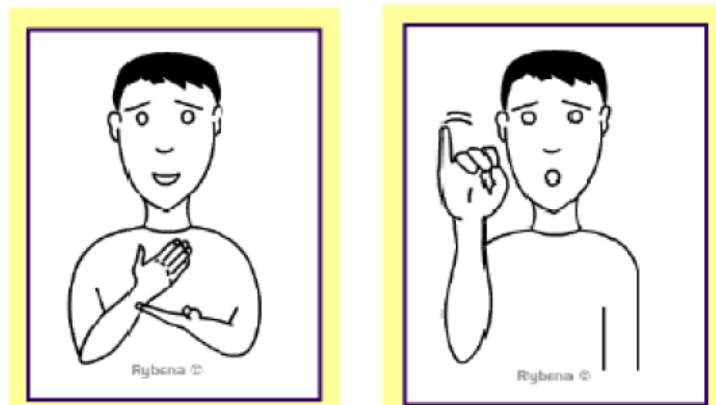
O *Rybená* [31] é um sistema que converte textos em língua portuguesa para LIBRAS. Ele foi desenvolvido em 2003 com o objetivo de incluir digitalmente brasileiros com deficiência auditiva. Primeiramente, foi lançada uma versão do sistema para ser executado em aparelhos de telefonia móvel com máquina virtual Java instalada, denominado de *Torpedo Rybená*. Posteriormente, foi desenvolvida uma versão dessa solução para Web, denominada *Player Rybená*.

O *Torpedo Rybená*, utilizado em aparelhos de telefonia móvel, permite que os deficientes auditivos se comuniquem com os ouvintes através de mensagens de texto, isto é, ele converte automaticamente mensagens de texto em língua portuguesa para LIBRAS. Os sinais em LIBRAS são representados por um agente animado virtual, denominado *Rybeninho*. Dessa forma, o deficiente auditivo pode receber, enviar e visualizar as mensagens em LIBRAS no seu celular. Na Figura 19 são ilustradas duas capturas de telas do *Torpedo Rybená*.

O *Player Rybená*, é uma adaptação da solução para navegação Web. Ele também é uma aplicação Java que converte textos de páginas Web em língua portuguesa para LIBRAS, utilizando o agente animado *Rybeninho*. Dessa forma, nos sites que utilizam o

Player Rybená, os deficientes auditivos podem selecionar textos da página Web e visualizar a tradução (conversão) em LIBRAS.

Figura 19: Torpedo Rybená [31].



Um dos problemas da solução *Rybená* é que ela utiliza tradução direta de português para LIBRAS, sem levar em consideração as diferenças gramaticais. Como a LIBRAS possui uma gramática própria, o processo de tradução direta, ou seja, sem utilização de processamento, interpretação ou análise de contexto, não é apropriado [20].

Uma adaptação da solução *Rybená* para ser aplicada em sistemas de televisão digital, denominada *RybenáTV*, foi proposta por Amorim *et al.* [32]. A ideia do *RybenáTV* é converter texto em português, extraído do *closed caption*, para LIBRAS, baseado na solução *Rybená*. Maiores detalhes podem ser encontrados em [32].

3.3. Considerações Finais

Neste capítulo descrevemos os principais trabalhos presentes na literatura científica relacionados à transmissão da língua de sinais e representação das informações não literais (efeitos sonoros e nuances emocionais). Os trabalhos citados, referente à representação dos sons produzidos nas cenas como também das emoções transmitidas pelo personagem, fornecem alternativas para a exibição dessas informações. Contudo, todos os trabalhos utilizam como base as línguas orais em formato texto, o que inviabiliza o entendimento do surdo, pois as línguas orais são

vistas apenas como uma segunda língua para os deficientes auditivos. Como citado anteriormente, a forma natural de comunicação entre os deficientes auditivos é através das línguas de sinais, portanto a forma mais indicada de exibição dessas informações é através da sua língua nativa.

Os trabalhos relacionados à transmissão da língua de sinais possuem taxas aceitáveis de transmissão e maneiras robustas de compressão e codificação, porém, as soluções citadas não dão suporte a exibição das informações não literais e diferenças regionais.

O presente trabalho propõe um protocolo que viabiliza a codificação e transmissão de janelas em língua de sinais para sistemas de televisão digital. Adicionalmente são sugeridas maneiras para representação das informações não literais. Como apenas um conjunto de códigos é transmitido e os vídeos da janela são montados no receptor de TV (ver capítulo 4), os aspectos regionais (regionalização) das línguas podem ser respeitados. As línguas de sinais, mais ainda que as línguas orais podem diferir muito entre uma região e outra de um mesmo país, desta forma, o caráter de regionalização se torna muito importante. Gerar o vídeo de sinais na estação de TV atende a apenas uma das muitas possíveis regionalizações existentes para uma determinada língua. Dada a dificuldade na troca de informações entre deficientes auditivos, em face da evidente falta de estrutura e tecnologia para tal, isto pode gerar um impacto decisivo para o seu entendimento. Os usuários poderiam, em uma solução baseada em dicionário, como proposta em [9], utilizar versões do dicionário adequadas a sua região. Além disso, essa solução utiliza pouca largura de banda do canal do radiodifusor para transmissão das janelas em línguas de sinais, uma vez que apenas um conjunto de códigos é transmitido.

No próximo capítulo serão apresentados os projetos relacionados à acessibilidade desenvolvidos no Laboratório de Aplicações de Vídeo Digital (LAVID). Os projetos buscam por novas dimensões do conceito de desenvolvimento inclusivo para promover a igualdade de oportunidades e direitos humanos, isto para que os portadores de deficiência auditiva sejam capazes de participar do processo comunicacional promovido pelas mídias digitais. Entre esses projetos está o LibrasTV, principal inspiração para criação do protocolo proposto nesse trabalho.

4. PROJETOS DE ACESSIBILIDADE

Neste capítulo é apresentada a descrição dos projetos de pesquisa desenvolvidos no Laboratório de Aplicações de Vídeo Digital (LAVID) relacionados à acessibilidade para deficientes auditivos. Dentre os projetos desenvolvidos estão: LibrasTV, que visa à geração automática das janelas de LIBRAS para sistemas de TV digital, GTAAS, que busca explorar o conceito de "acessibilidade como um serviço", propondo um serviço de computação em nuvem para ajudar as pessoas surdas a acessar conteúdos digitais, e o projeto CineLIBRAS, que tem como objetivo a geração automática de legendas em LIBRAS para cinemas digitais.

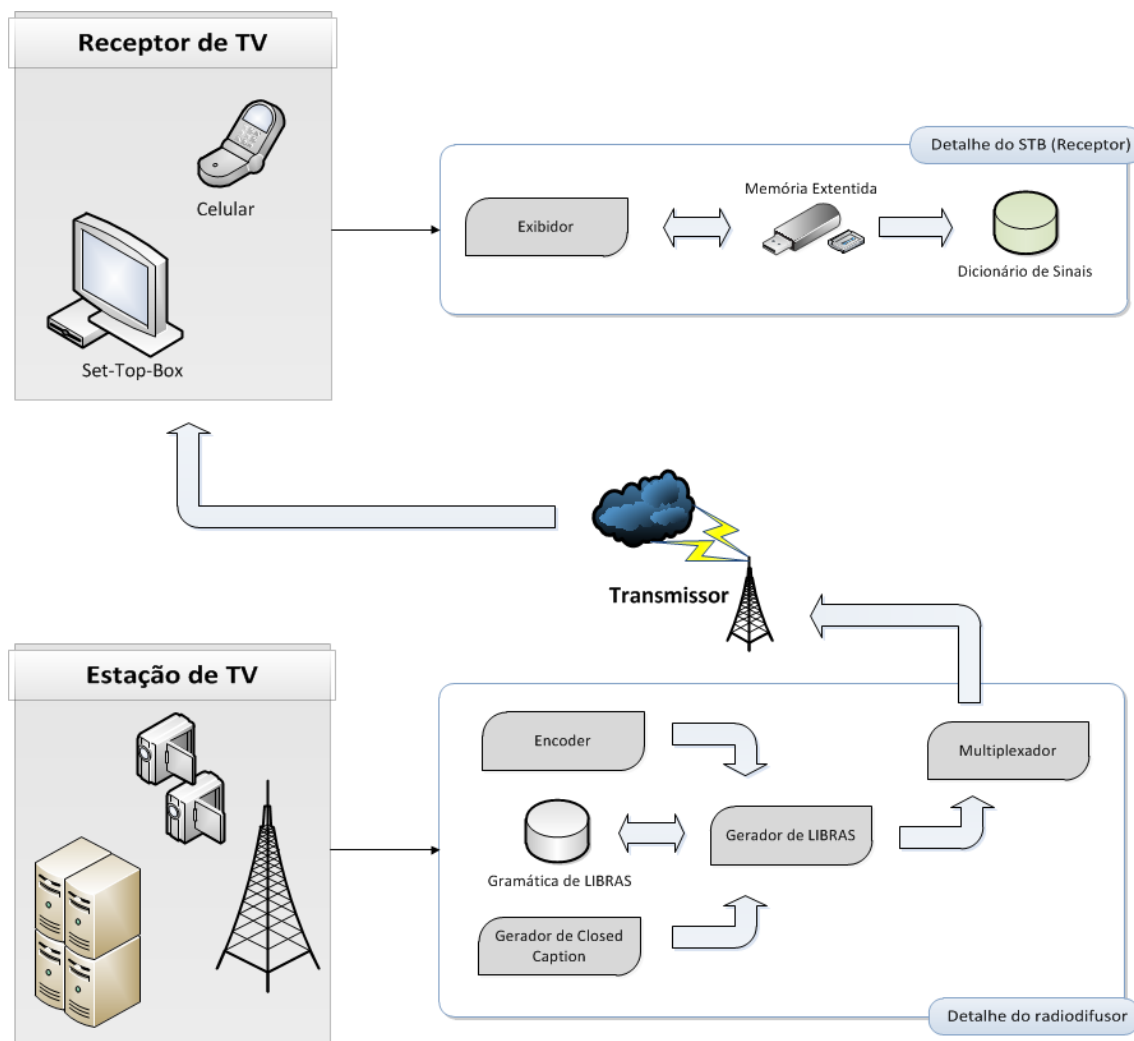
4.1. Projeto LibrasTV

Nesta seção descreveremos brevemente a arquitetura do projeto LibrasTV. Nessa solução, um conjunto de componentes permite a geração automática da janela de LIBRAS (ou seja, a tradução sem um intérprete humano) a partir de um fluxo de entrada de áudio ou *closed caption* em PB. A arquitetura de funcionamento do sistema segue a proposta de Araújo *et. al.* [9]. A Figura 20 exemplifica a arquitetura proposta.

O LibrasTV é um projeto de pesquisa e desenvolvimento que se propõe a desenvolver uma solução para suporte a LIBRAS no Sistema Brasileiro de TV Digital. Neste projeto são desenvolvidos um doutorado (relacionado à parte de tradução automática) e três mestrados acadêmicos (um referente à geração da LIBRAS codificada e dois na parte de recepção e exibição), além da participação de sete alunos de graduação. Alunos de graduação em Ciência da Computação, Pedagogia e uma aluna deficiente auditiva fazem parte do corpo permanente do projeto. O objetivo é criar e integrar componentes de software e hardware que funcionarão em dois módulos distintos. O primeiro desses módulos funcionará no radiodifusor, e tem o objetivo de gerar, codificar e transmitir legendas em LIBRAS em um fluxo de TV Digital, associada a um serviço existente. O segundo módulo, se propõe a decodificar, sincronizar e apresentar as legendas em LIBRAS na estação receptora.

Uma característica importante deste projeto é a utilização de um dicionário para armazenar representações visuais de sinais em LIBRAS. No dicionário de LIBRAS, cada sinal pode ser representado por uma imagem animada ou um arquivo de vídeo (por exemplo, um GIF ou um arquivo de vídeo) com um código (por exemplo, uma representação textual) associado a sua representação. O projeto também define uma estratégia para sincronizar o fluxo de entrada de áudio (ou texto) e o fluxo de saída de vídeo de LIBRAS.

Figura 20: Arquitetura proposta para a transmissão de LIBRAS no Sistema Brasileiro de TV Digital.



Inicialmente, o componente Tradutor de LIBRAS localizado no Gerador de LIBRAS recebe um fluxo de entrada de áudio ou *closed caption*. Um processo de reconhecimento de voz ou de extração de *closed caption* é então aplicado para converter esse fluxo em uma sequência de palavras em formato de texto. Em seguida, a sequência

de palavras é automaticamente traduzida para uma sequência de glosas (representação textual em LIBRAS) que são codificados com informações de sincronização e transmitidos em um canal de comunicação. O fluxo gerado pelo processo de codificação é chamado fluxo codificado de LIBRAS. Mais detalhes sobre o processo de codificação são apresentados no capítulo 5.

Finalmente, o componente exibidor, localizado no receptor, recebe o fluxo codificado de LIBRAS a partir do canal de comunicação, decodifica, sincroniza e exibe os sinais para gerar a janela de LIBRAS. É composto por dois processos: o de decodificação e exibição. O processo de decodificação extrai a sequência de glosas e informações de sincronização do fluxo de LIBRAS codificado e o processo de exibição associa cada glosa com a sua representação visual armazenada no Dicionário de LIBRAS. Assim, a sequência de glosas é convertida para uma sequência de representações visuais, que será sincronizada para gerar a janela LIBRAS.

A sincronização entre o fluxo de entrada (áudio ou *closed caption*) e o fluxo de saída (janela de LIBRAS) é realizada usando eventos de sincronismo DSM-CC, como visto anteriormente. Este modelo define pontos de sincronização que são inseridos no fluxo usando timestamps com base em um cronômetro global. Neste caso, o temporizador global é o relógio referido no fluxo de entrada (áudio ou *closed caption*). Este relógio é utilizado para gerar os timestamps que serão inseridos no fluxo de saída (isto é, na janela de LIBRAS).

Nas próximas subseções iremos detalhar os principais componentes do projeto LibrasTV.

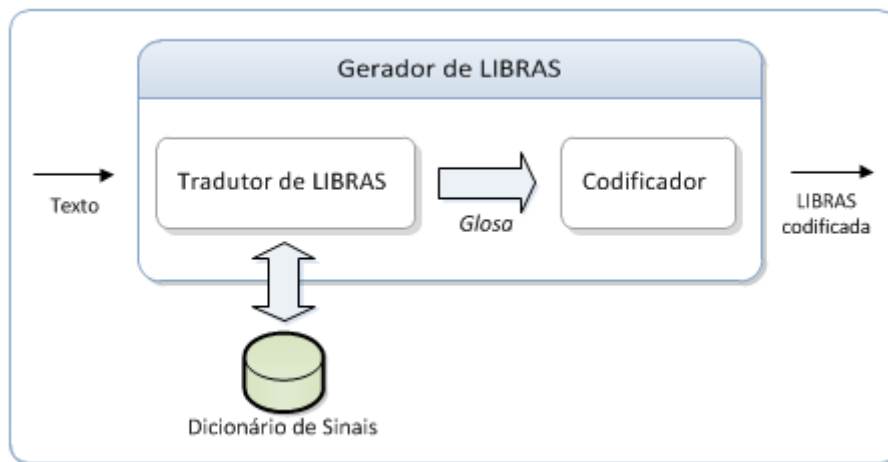
4.1.1. Gerador de LIBRAS

O Gerador de LIBRAS é o componente responsável por receber o texto em língua portuguesa, traduzir para a forma textual da língua brasileira de sinais (glosa) e codificá-la. Ele é composto pelos componentes Tradutor de LIBRAS e Codificador.

Como mencionado anteriormente, o Tradutor de LIBRAS é responsável por traduzir o fluxo de entrada (ou seja, áudio ou *closed caption*) para uma representação textual em LIBRAS (sequência de glosas) para posterior codificação e transmissão em

um canal de comunicação. O reconhecimento de voz e o processo de extração do *closed caption* são usados para converter o fluxo de entrada (áudio ou *closed caption*, respectivamente) numa sequência de palavras em língua portuguesa. A tradução automática, os processos de codificação e exibição serão detalhados mais a frente. A Figura 21 exibe o componente Gerador de LIBRAS de forma mais detalhada.

Figura 21: Detalhamento do Gerador de LIBRAS.



a) Processo de tradução automática

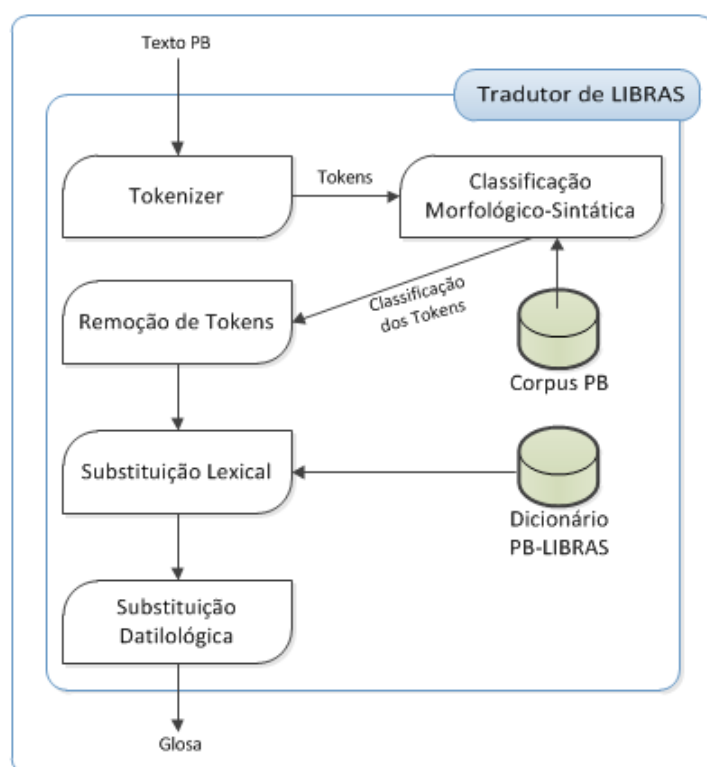
O processo de tradução automática é realizado pelo componente Tradutor de LIBRAS para converter uma representação textual em PB para uma representação textual (sequência de glosas) em LIBRAS. Este processo baseia-se nos passos ilustrados na Figura 22.

No primeiro passo (ou seja, o passo Tokenizer), o texto no PB é dividido em uma sequência de palavras (ou tokens). Posteriormente, os tokens são classificados em categorias morfológico-sintáticas. Para fazer esta tarefa, foi usado o PPM-C [33], uma variante de um algoritmo de compressão de dados estatísticos adaptativos baseados em modelos ocultos de Markov (HMM) chamados Previsão por Correspondência Parcial (PPM) [34], devido à sua capacidade de construir modelos estatísticos precisos [35] e ter sido usado em outros problemas de classificação [36] [37] [38].

O PPM constrói um modelo estatístico e usa-o para armazenar a frequência de diferentes sequências de elementos. Depois que o modelo é construído, o próximo

elemento da sequência pode ser previsto de acordo com seus 'n' elementos anteriores. Como os valores mais próximos de 'n' tendem a aumentar o tempo necessário para calcular o algoritmo, a ordem do algoritmo deve ser escolhida previamente, mantendo os tempos de conclusão aceitáveis. A variante PPM-C é mais eficiente que a implementação original em termos de tempo de execução e espaço de dados, em troca de uma compressão ligeiramente inferior.

Figura 22: Processo de tradução automática.



A classificação morfológico-sintática utiliza classes morfológico-sintáticas como elementos no PPM-C. O modelo construído para ele armazena sequências das classes citadas anteriormente tomadas a partir de um corpus com classificação morfológico-sintática em PB. Uma vez que uma frase é recebida para classificação, a classe morfológica mais provável de cada símbolo é selecionada de acordo com o seu contexto, em termos do modelo de construção.

Depois de classificar os tokens, são aplicadas algumas regras de tradução (definidas por especialistas humanos) para traduzir esses tokens (ou palavras) para a representação em glosa. Inicialmente, o texto é simplificado, eliminando alguns tokens

(passo Remoção de Tokens). Esse passo é executado pelo fato da LIBRAS não utilizar preposições e artigos. Assim, este conjunto de tokens pode ser removido. Depois, alguns tokens (ou palavras) são substituídos (substituição lexical), a fim de adaptar o significado da frase reescrita em LIBRAS. O vocabulário de LIBRAS é consideravelmente menor do que o PB [39]. Por exemplo, as palavras lar, casa, habitação em PB têm o mesmo sinal (ou seja, a mesma representação visual) em LIBRAS. Além disso, embora os verbos em PB tenham um elevado grau de flexão os verbos em LIBRAS não são flexionáveis. Em seguida, os verbos em PB são substituídos por verbos em glosa não flexionáveis (ou seja, os verbos de LIBRAS). Para fazer essa substituição, usamos um conjunto de sinônimos de PB para LIBRAS (Dicionário PB-LIBRAS) desenvolvido por especialistas em LIBRAS. Finalmente, os nomes próprios e termos técnicos são soletrados em LIBRAS (por formas de mão que representam as letras da palavra). A saída de todo este processo é a representação da LIBRAS em forma textual (glosa).

b) Codificador

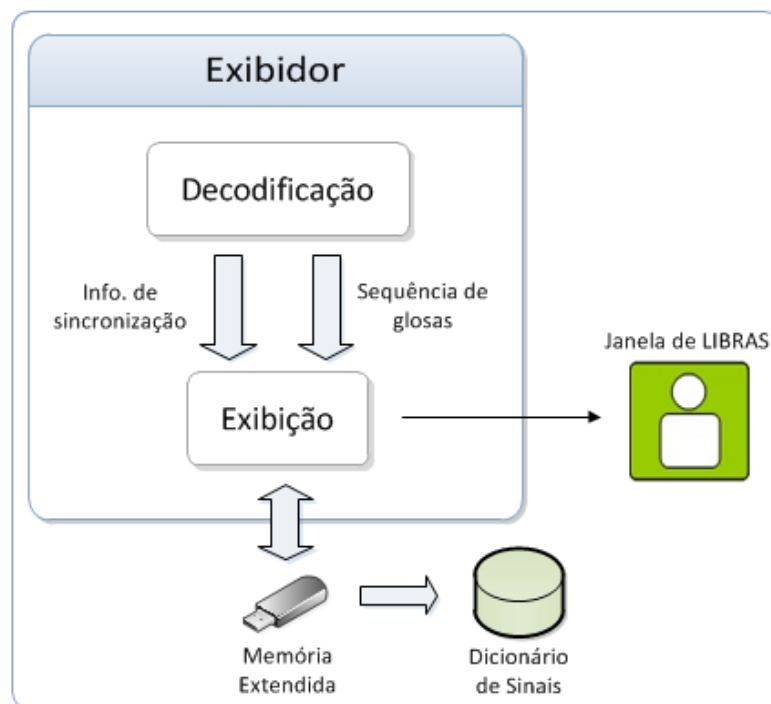
O módulo codificador é responsável pela codificação da sequência de glosas gerada no Tradutor de LIBRAS com as informações de sincronização gerada nos módulos de sincronização. A saída deste componente é o fluxo de LIBRAS codificado (*elementary stream* – ES). Esse ES é composto por um conjunto de códigos, onde cada código está relacionado a uma animação armazenada em um Dicionário de LIBRAS localizado no receptor. 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. O ES é utilizado pelo componente exibidor, para exibir e sincronizar os sinais e, portanto, para gerar a janela LIBRAS.

Uma característica importante da solução 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 nesta ideia, 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.

4.1.2. Exibidor

O componente exibidor, localizado no receptor, é responsável por extrair os dados (sequência de glosas e informações de sincronização) a partir do fluxo de LIBRAS codificado. Também é responsável por decodificar e exibir os sinais de forma síncrona. De acordo com a Figura 23, o exibidor é composto por dois módulos principais: decodificação e exibição. O módulo de decodificação recebe o fluxo de LIBRAS codificado e extrai a sequência de glosas e as informações de sincronização deste fluxo. O módulo de exibição recebe a sequência de glosas, associa cada glosa (sinal) com a sua representação visual no Dicionário de LIBRAS, armazenado em um dispositivo de memória estendida, como, por exemplo, um dispositivo de armazenamento USB, e exibe a representação visual dos sinais. Para sincronizar os sinais com o *closed caption*, o componente Exibidor usa as informações de sincronização, citadas anteriormente.

Figura 23: Detalhamento do módulo exibidor.



Assim, a sequência de glosas é convertida em uma sequência de representações visuais que são sincronizadas para gerar a janela LIBRAS. Também é definida uma

configuração neutra (posição, cor de fundo, brilho) para começar e terminar a representação de cada sinal. Assim, o módulo de exibição pode suavizar a transição entre sinais consecutivos. Este componente também pode oferecer funcionalidades adicionais, tais como ativar / desativar, redimensionar e reposicionar a janela de LIBRAS. Este recurso permite aos usuários escolher suas configurações de exibição.

4.1.3. LibrasTV em sistemas digitais

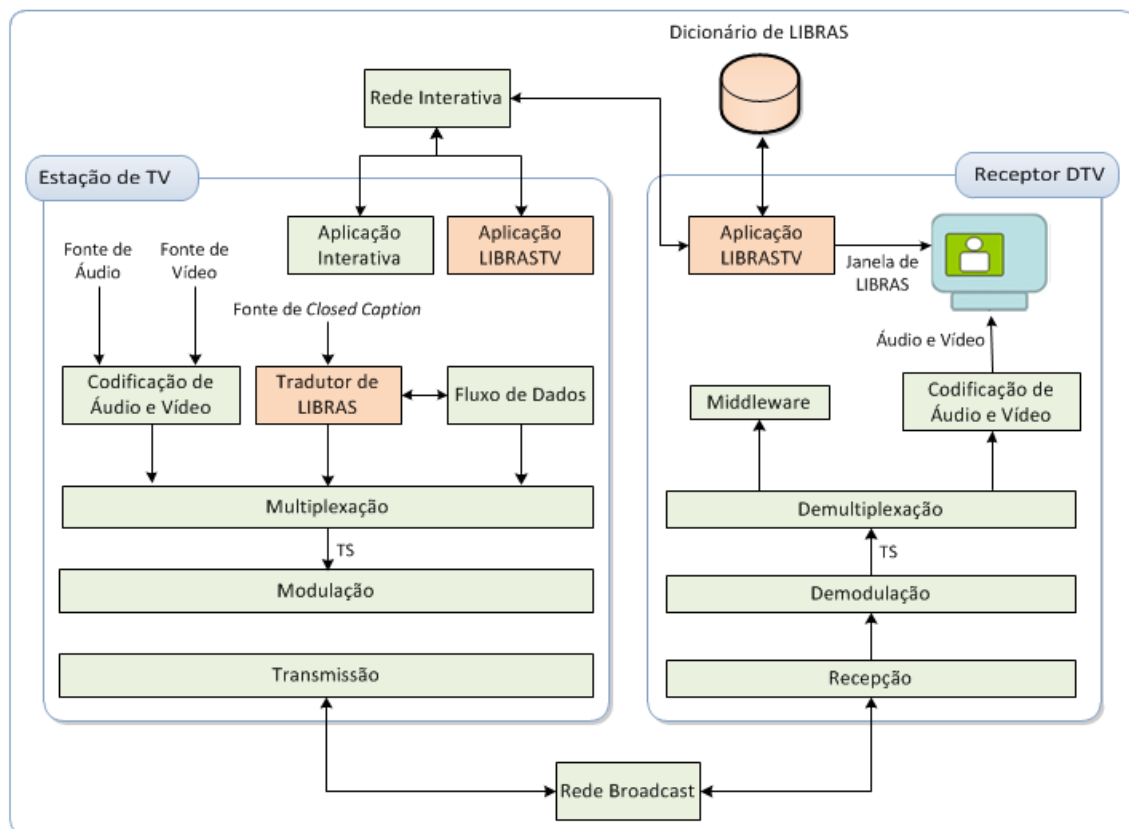
Existem outras maneiras possíveis de integração do LibrasTV em sistemas de TV digital, como por exemplo: (1) integração de todos os componentes na estação de TV (a janela de LIBRAS seria gerada na estação de TV e transmitida em MPEG-2 TS como um vídeo secundário) e (2) integração de todos os componentes nos receptores de TV digital. A primeira solução consome muita largura de banda, uma vez que um vídeo secundário é transmitido em MPEG-2 TS, e não preserva as diferenças regionais, pois será gerada uma mesma janela de LIBRAS para todas as regiões. A segunda solução também tem algumas limitações, pois exige muito processamento em receptores de TV digital para traduzir, codificar, decodificar, sincronizar e exibir a janela de LIBRAS. Estes argumentos justificam a decisão de integrar o LibrasTV em sistemas digitais utilizando a abordagem mostrada com mais detalhes na Figura 24.

Um sistema de TV digital é basicamente um sistema cliente-servidor onde o servidor é a estação de televisão (ou provedor de conteúdo) e o cliente é o ambiente do usuário (como visto na Figura 24). Na estação de TV, as fontes de vídeo e áudio analógicos (capturados a partir de uma câmera ou a partir de um servidor de vídeo) são entregues a codificadores digitais, que são responsáveis pela codificação e compressão dos fluxos de áudio e vídeo.

Então, esses fluxos de áudio e vídeo são comprimidos e multiplexados em conjunto com fluxos de dados em um único fluxo, chamado fluxo de transporte (MPEG-2 TS - Transport Stream). O MPEG-2 TS é então modulado e transmitido em uma rede de transmissão (por exemplo, terrestre, cabo, satélite). No lado do receptor, o sinal é recebido, demodulado e entregue ao demultiplexador, que separa os fluxos de vídeo, áudio e de dados. Os fluxos de áudio e vídeo são enviados para os decodificadores, que decodificam e sincronizam os sinais para a exibição, enquanto os

fluxos de dados são enviados para serem processados pelo middleware. A aplicação interativa pode também exigir novos dados que podem ser obtidos a partir do canal interativo (ou de retorno).

Figura 24: LibrasTV aplicada em sistemas de TV digital (componentes LibrasTV estão em destaque)



O LibrasTV também poderia ser incluído em outras plataformas, tais como a Web e em Cinemas Digitais, como veremos nas seções seguintes.

4.2. Grupo de Trabalho Acessibilidade como um Serviço (GTAaaS)

Nesta seção descrevemos brevemente o projeto GTAaaS (ver Figura 25). O serviço proposto neste projeto de pesquisa explora "acessibilidade como um serviço" gerando automaticamente e incorporando uma faixa de vídeo em língua de sinais em conteúdos multimídia. O serviço conta com a colaboração de especialistas em língua de

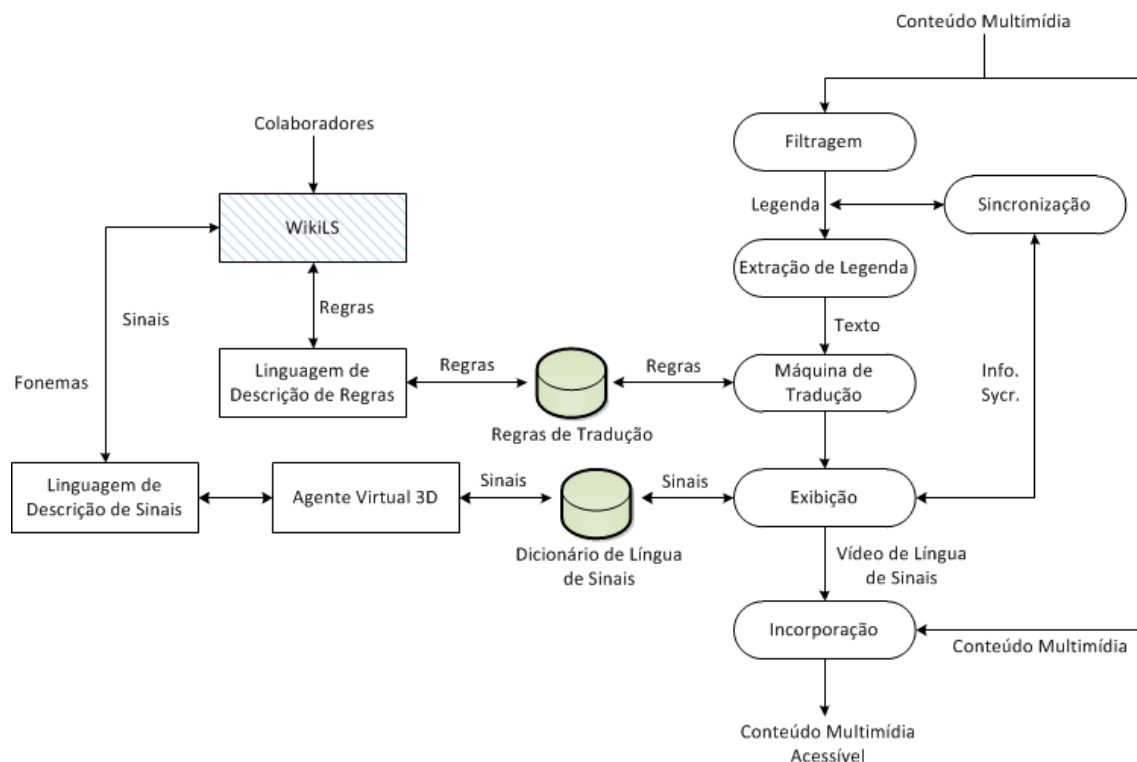
sinais para ajustar dinamicamente o sistema que roda em uma infraestrutura de computação em nuvem.

Um aspecto importante da proposta é a utilização de estratégias de colaboração para melhorar / evoluir a qualidade da tradução e apresentação do serviço ao longo do tempo. A ideia é que especialistas em línguas de sinais possam ajudar a melhorar o serviço, por exemplo, através da melhoria das regras de tradução, incluindo novos sinais. Para fazer esta tarefa, foram definidas linguagens que descrevem regras de tradução (a linguagem de descrição de regras) e sinais (a linguagem de descrição de sinais), e uma ferramenta de colaboração para lidar com elas (a ferramenta WikiSL).

De acordo com a Figura 25, o serviço funciona da seguinte forma. Inicialmente, o conteúdo multimídia submetido é aplicado a um componente de filtragem, que extrai a faixa de legendas. Depois, um componente de extração de legendas converte este fluxo de legendas em uma sequência de palavras na língua falada fonte. Em seguida, esta sequência de palavras é automaticamente traduzida (usando o componente Máquina de Tradução, similar ao Tradutor de LIBRAS do projeto LibrasTV) para uma sequência de palavras na língua de sinais alvo (isto é, uma sequência de glosas), de acordo com um conjunto de regras de tradução. A sequência de glosas é então enviada para um componente de exibição que associa cada glosa com uma representação visual de um sinal armazenado em um dicionário de língua de sinais. Assim, a sequência de glosas é mapeada para uma sequência de representações visuais que serão sincronizadas com a faixa de legendas para gerar o vídeo de língua de sinais. Finalmente, o vídeo de língua de sinais está embutido no conteúdo multimídia como uma camada de vídeo extra (pela incorporação de componentes), tornando-o acessível para os surdos.

Sincronização entre o conteúdo regular multimídia e o vídeo de língua de sinais é realizada com base em eixos baseados em modelos de sincronização [40]. São definidos pontos de sincronização e posteriormente inseridos nos conteúdos multimídias usando timestamps com base em um temporizador global. Nessa solução, o temporizador global é o relógio de referência da faixa de legendas. Este relógio é extraído a partir da faixa de legendas e é utilizado para gerar os timestamps de apresentação para os sinais do vídeo de língua de sinais.

Figura 25: Visão esquemática do serviço proposto.



4.3. CineLIBRAS

Nesta seção descreveremos as principais características do projeto. O CineLIBRAS preconiza a geração automática das legendas (janela) em Língua Brasileira de Sinais (LIBRAS) para a utilização em cinemas digitais. O Cinema Digital surgiu diante do contínuo avanço das pesquisas em torno da produção, armazenamento, distribuição e apresentação de filmes digitais em *Ultra High Definition* - UHD. Um dos esforços resultantes dessas pesquisas foi a definição e padronização de um pacote para transmissão dos dados, denominado *Digital Cinema Package* - DCP. O DCP utiliza todas as mídias digitais no processo de multiplexação, como, por exemplo, imagem, áudio, legenda e dados, organizando suas trilhas em formato de uma lista de exibição (*playlist*) e direcionado aos equipamentos de apresentação.

Nessa solução a produção das legendas em LS é feita na própria sala de cinema, sem a utilização de um intérprete de Língua de Sinais. Para atingir este objetivo, foi desenvolvido um dicionário de LIBRAS, com sinais representados por um agente virtual animado 3D (avatar-3D). A legenda gerada pode ser apresentada em dispositivos de *displays* externos, tais como *tablets*, *smartphones* e similares, que possuam recursos

para o recebimento e a apresentação de fluxos de vídeo. A possibilidade de apresentação em dispositivos móveis ou remotos se deve ao fato do incômodo proporcionado pela visualização de uma janela de LIBRAS sobrepondo o vídeo principal nas telas de cinema.

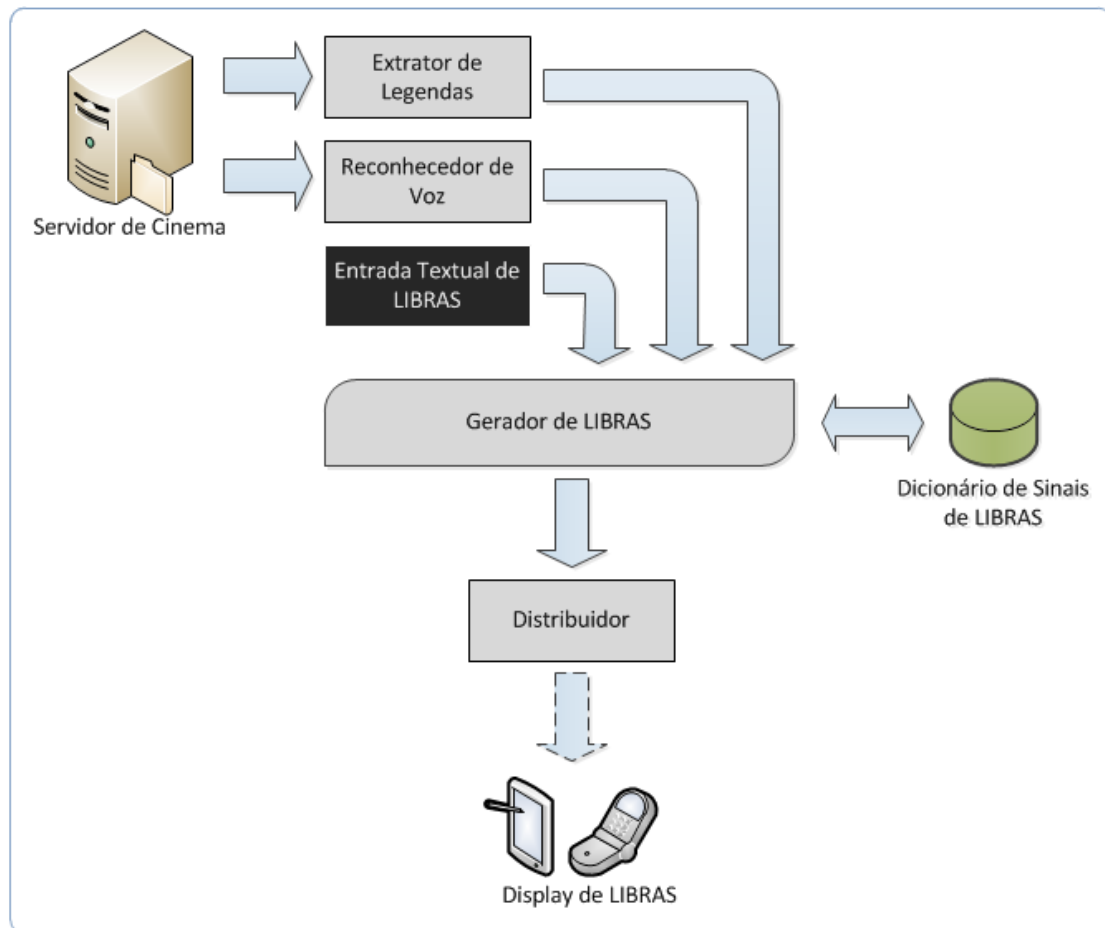
A geração da janela na própria sala de cinema diminui o impacto de mudança no processo de exibição do filme, já que não gera conflitos nos processos de produção e distribuição, além de contemplar a questão da regionalização (representação visual de um sinal conhecida em determinada área que adota uma língua de sinais). Para isto, o sistema faz uso de um dicionário de LIBRAS, que deve conter a representação visual dos sinais de LIBRAS de acordo com a região, isto é, considerando os dialetos regionais da LIBRAS, e respeitando o seu caráter de língua natural. Como citado anteriormente, por exemplo, um sinal de LIBRAS no Rio de Janeiro pode ser representado diferentemente em São Paulo. Em consequência disso, a geração das legendas em LIBRAS antes da fase de distribuição dos filmes poderia resultar, por exemplo, na produção da janela para uma das regiões, dificultando o entendimento do filme por deficientes auditivos localizados em outras regiões de um mesmo país.

Os componentes para a geração automática de legendas em LIBRAS para Cinemas Digitais (Figura 26) podem ser resumidos da seguinte forma:

- *Reconhecedor de Voz* – processa o sinal de áudio contido nos arquivos referentes à trilha de áudio do filme e gera uma saída de texto, e pode ser utilizado como um dos possíveis sinais de entrada do Gerador de LIBRAS. A trilha de legendas é enviada ao *Reconhecedor de Voz* por um Player de Cinema / Servidor de Cinema Digital.
- *Extrator de Legendas* – recebe do Player de Cinema / Servidor de Cinema Digital os arquivos (extensão ‘xml’) referentes à trilha de legendas do filme, extrai o texto falado e é um dos possíveis sinais de entrada do Gerador de LIBRAS.
- *Entrada Textual de LIBRAS* – é outro possível sinal de entrada do Gerador de LIBRAS. Esse sinal pode ser gerado, por exemplo, por um operador de LIBRAS.
- *Player de Cinema/Servidor de Cinema Digital* – é um *player* de cinema, como os utilizados atualmente, responsáveis por selecionar e encaminhar as

trilhas de legenda e/ou de áudio para processamento no *Extrator de Legendas* e no *Reconhecedor de Voz*, respectivamente.

Figura 26: Arquitetura do sistema para geração automática de legendas em LIBRAS para cinemas digitais.



- *Gerador de LIBRAS* – recebe como entrada um sinal de texto e, após aplicar processo de tradução entre a língua falada gera, com o auxílio do Dicionário de Sinais de LIBRAS, um fluxo (*stream*) de vídeo que contém as legendas (janela) em LIBRAS. A associação entre o texto e o sinal de LIBRAS é feita através do Dicionário de Sinais.
- *Distribuidor* – distribui o fluxo (*stream*) de vídeo contendo legendas em LIBRAS para dispositivos de visualização utilizando o protocolo proposto no capítulo 5 e uma infraestrutura de transmissão de dados.
- *Display de LIBRAS* – é um dispositivo de visualização que recebe um fluxo (*stream*) de vídeo contendo legendas em LIBRAS e as apresenta para o usuário.

4.4. Considerações Finais

Neste capítulo descrevemos os projetos desenvolvidos no Laboratório de Aplicações de Vídeo Digital (LAVID) que buscam a inclusão social do surdo no processo comunicacional visando promover os deficientes auditivos a membros de audiência. Foram descritos os principais componentes dos projetos de pesquisa LibrasTV (principal motivação para criação do protocolo proposto neste trabalho), GTAAaS e CineLIBRAS. O projeto LibrasTV visa à geração automática das janelas de LIBRAS a partir de um fluxo de entrada de áudio ou *closed caption* em Português Brasileiro (PB) para sistemas de TV digital, o projeto GTAAaS gera automaticamente e incorpora uma faixa de vídeo em língua de sinais em conteúdos multimídias, já o CineLIBRAS tem como objetivo a geração automática de legendas em LIBRAS para cinemas digitais, proporcionando o acesso dos deficientes auditivos a salas de cinema. No próximo capítulo iremos apresentar as principais informações sobre o protocolo para transmissão da Língua Brasileira de Sinais (LIBRAS), foco deste trabalho.

5. PROTOCOLO PROPOSTO

Nesse capítulo é apresentada a proposta de protocolo de codificação de janelas de LIBRAS. Nessa proposta, os quadros do vídeo da janela de LIBRAS não são transmitidos. Em vez disso, um conjunto de códigos é enviado, no qual cada código está associado à representação visual de um sinal de LIBRAS, armazenada no dicionário de LIBRAS do receptor (ver Capítulo 4). Conforme mencionado no Capítulo 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, os dicionários de LIBRAS podem estar armazenados nos clientes, através de dispositivos (USB – *Universal Serial Bus*) ou serem carregados pelo canal de comunicação, permitindo que as diferenças regionais presentes nas línguas de sinais sejam respeitadas, adaptando a apresentação da janela de LIBRAS ao seu contexto, pois alguns sinais podem ser representados diferentemente de acordo com a região do país. Por exemplo, alguns sinais em LIBRAS podem ser representados diferentemente no nordeste, sul e sudeste do Brasil. Além disso, o protocolo também permite que informações associadas às emoções, ao humor e efeitos sonoros sejam codificadas e transmitidas.

O protocolo define dois tipos de mensagens, a *librasControlMessage* e a *librasDataMessage*, e são identificadas a partir dos valores contidos na Tabela 1. As mensagens de controle, *librasControlMessage* (LCM, ver Tabela 2), 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.

Tabela 1: Valores possíveis para o campo `message_id`.

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

Tabela 2: Estrutura da *librasControlMessage*.

librasControlMessage () {	
message_id	8 bits;
message_length	16 bits;
sign_language_content_type	8 bits;
network_id	16 bits;
transport_stream_id	16 bits;
service_id	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;
}	
}	

Os campos da estrutura *librasControlMessage* possuem os seguintes significados:

message_id - É um campo de 8 bits que identifica o tipo de mensagem de LIBRAS. Deve ser configurado com o valor 0x01, conforme a Tabela 1;

message_length - É um campo de 16 bits que representa o número de bytes da mensagem, imediatamente após o campo *message_length*, e incluindo o CRC. Esse campo não pode ultrapassar o valor 65533;

sign_language_content_type – É um campo de 8 bits utilizado para configurar o tipo de conteúdo de língua de sinais que está sendo transmitido. Ele carrega o valor ‘1’ quando existe um stream de vídeo com legendagem em língua de sinais executada por um intérprete humano sendo transmitido. Quando o valor do campo *sign_language_content_type* é ‘1’, não existe nenhuma *SLDataMessage* sendo transmitida. O valor ‘2’ indica que a aplicação deve monitorar o recebimento de *SLDataMessages* e, usando o conteúdo transmitido neste tipo de mensagem, gerar automaticamente o conteúdo de língua de sinais utilizando um dicionário de sinais;

network_id – É um campo de 16 bits que deve obrigatoriamente conter um valor único de identificação da rede;

resolução - É um campo de 4 bits que representa a resolução inicial da camada gráfica de apresentação de LIBRAS no receptor. Os possíveis valores do campo resolução são apresentados na Tabela 3;

transport_stream_id: campo de 16 bits que deve obrigatoriamente servir como um rótulo para identificar o TS, diferenciando-o de qualquer outro multiplexador dentro do sistema de transmissão;

service_id: campo de 16 bits que deve obrigatoriamente indicar o service_id do evento, que anuncia dados originais e parciais do programa. O service_id deve obrigatoriamente ser igual ao program_number dentro do program_map_section correspondente;

Tabela 3: Campo resolução.

Valor	Resolução
0x00	1920x1080
0x01	1280x720
0x02	640x480
0x03	960x540
0x04	720x480
0x05-0x0F	reservado

posição_x - É um campo de 16 bits que representa a coordenada horizontal da janela de exibição de LIBRAS, a partir do canto superior esquerdo da tela, em relação à resolução da camada gráfica de LIBRAS;

posição_y - É um campo de 16 bits que representa a coordenada vertical da janela de exibição de LIBRAS, a partir do canto superior esquerdo da tela, em relação à resolução da camada gráfica de LIBRAS;

largura - É um campo de 16 bits que representa o comprimento (em número de pontos) da janela de exibição de LIBRAS, em relação à resolução da camada gráfica de LIBRAS;

altura - É um campo de 16 bits que representa a altura (em número de pontos) da janela de exibição de LIBRAS, em relação à resolução da camada gráfica de LIBRAS;

private_data_length - É um campo de 8 bits que representa o número de bytes de dados privados que seguem imediatamente o campo private_data_length;

private_data_bytes - É um campo definido pelo usuário.

O protocolo também define as mensagens de dados, *librasDataMessage* (LDM, ver Tabela 4), que transmitem o conjunto de códigos dos sinais de LIBRAS e as informações associadas às emoções, ao humor e aos efeitos sonoros presentes no *closed caption* como por exemplo, a, entonação de voz e nuances emocionais, etc. Além disso, na LDM também é possível identificar o personagem que está discursando.

Para encapsular as mensagens LCM e LDM no fluxo MPEG-2 TS (protocolo de transporte adotado por todos os sistemas de TV atuais [13] uma alternativa é utilizar os eventos definidos na especificação *Digital Storage Media-Command and Control* (DSM-CC) [12], denominados eventos DSM-CC, apresentados na Seção 2.2 e discutidos na Seção 5.5.

Tabela 4: 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	16 bits;
for (i = 0; i < N; i++) {	
emotions_flag	1 bit;
reservado	7 bits;
glosa_data_bytes	8 bits;
}	
}	
}	
}	

Os campos da estrutura *librasDataMessage* possuem os seguintes significados:

message_id - É um campo de 8 bits que identifica o tipo de mensagem de LIBRAS. Deve ser configurado com o valor 0x02, conforme a Tabela 1;

message_length - É um campo de 16 bits que representa o número de bytes da mensagem, imediatamente após o campo `message_length`, e incluindo o CRC. Esse campo não pode ultrapassar o valor 65533;

identifier_control_flag - É um campo de 1 bit que indica que deve ser armazenado o nome do personagem falante da cena;

identifier_data_length - É um campo de 8 bits que representa o número de bytes do identificador do personagem;

identifier_data_bytes - Representação em bytes do identificador do personagem falante da cena;

parameters_control_flag - É um campo de 1 bit que indica que devem ser armazenados parâmetros não funcionais como efeitos sonoros e nuances emocionais;

parameters_loop_length - É um campo de 8 bits que representa o número de bytes de representação dos efeitos sonoros e nuances emocionais;

parameters_data_length - É um campo de 8 bits que representa o número de bytes dos parâmetros não funcionais;

parameters_data_bytes - Representação em bytes dos parâmetros não funcionais;

sinais_loop_length - É um campo de 16 bits que representa o número de bytes de representação dos sinais seguindo imediatamente o campo `sinais_loop_length`;

glosa_data_length - É um campo de 8 bits que representa o número de bytes da representação em *glosa*³ do sinal;

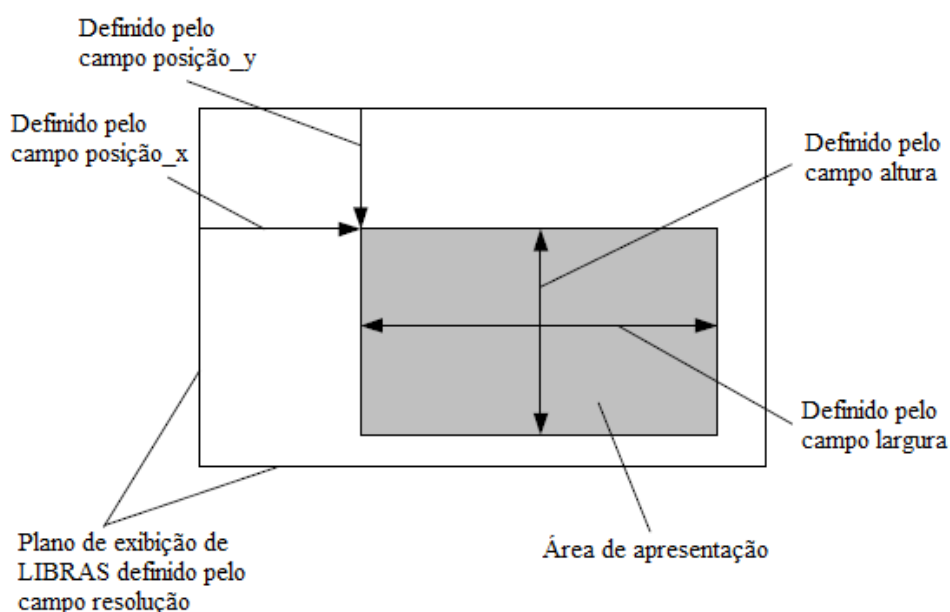
emotions_flag - É um campo de 1 bit que indica em que ponto devem ser exibidas as expressões faciais referente a emoções, interjeições, indagações, ironias e outros parâmetros não funcionais. Se este campo tiver valor '0', a expressão facial referente à emoção não será exibida. O valor '1' indica que a expressão facial relacionada a emoção deve ser exibida;

glosa_data_bytes - Representação em bytes da *glosa* do sinal.

³ Um texto em estrutura gramatical seguindo a estrutura da Língua Portuguesa para um texto com uma estrutura semelhante à utilizada pela Língua Brasileira de Sinais, essa passagem de um estado para o outro é comumente denominado *glosa*.

A sintaxe das mensagens LCM e LDM é apresentada nas Tabelas 2 e 4, respectivamente. Conforme ilustrado nessas tabelas, as duas mensagens sempre iniciam com um campo **message_id**, que identifica o tipo da mensagem (ver Tabela 1). Conforme pode ser observado na Tabela 2, 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 (ver Figura 27).

Figura 27: Plano de exibição e área de apresentação de LIBRAS.



A mensagem LDM (ver Tabela 4) é composta principalmente pelos campos: *identifier_data_bytes*, *parameters_data_bytes* e *glosa_data_bytes*. O campo *identifier_data_bytes* é utilizado para identificar o personagem que está falando. Por exemplo, se o personagem Josué está discursando na cena, o valor JOSUE será representado no campo *identifier_data_bytes*, de forma similar ao formato utilizado para identificar o personagem em legendas ocultas (ver Figura 28).

Efeitos sonoros e as emoções expressadas nas falas dos personagens presentes nas legendas ocultas (*closed caption*) podem ser fatores determinantes para a

compreensão das informações transmitidas nas cenas. A presença dessas informações na sinalização são semanticamente importantes e sua ausência pode comprometer o entendimento dos conteúdos apresentados. Pensando nesse problema o presente trabalho sugere maneiras para codificação e apresentação dessas informações.

Como podemos ver na Figura 28, a emoção (desesperado) transmitida pelo personagem Josué e os efeitos sonoros (Spray, Freada, Estrondo e Tiro) presentes nas imagens da Figura 29 são denominados informações não literais e seus valores são armazenados no campo *parameters_data_bytes*.

Figura 28: Cena do filme *Central do Brasil* com *closed caption*, demonstrando a identificação do personagem e a emoção transmitida.



Figura 29: Cenas de filmes variados, com *closed caption*, demonstrando efeitos sonoros.



Na Tabela 5 são apresentados alguns possíveis valores para esse campo.

Tabela 5: 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.

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, e codificados como um conjunto de glosas, em que cada glosa representa uma letra do nome.

5.1. Sincronização e Encapsulamento

Outro requisito importante do protocolo de codificação é definir qual a estrutura usada para encapsular as suas mensagens em um fluxo de TV digital, MPEG-2 *Transport Stream* (MPEG-2 TS) (conforme visto no Capítulo 2, Seção 2.3) e como apresentar os sinais sincronizadamente com o vídeo do programa.

Uma possível solução para os dois problemas é transportar as mensagens no protocolo de codificação de eventos definido pela especificação *Digital Storage Media-Command and Control* [12], os eventos DSM-CC, (conforme visto no Capítulo 2, Seção 2.4).

Os eventos DSM-CC, transmitidos em estruturas denominadas *Stream Event Descriptor* (ver Tabela 6), 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, é utilizada para transportar as mensagens do LCM e LDM do protocolo proposto.

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

Tabela 6: Estrutura do *Stream Event Descriptor*.

StreamEventDescriptor () {	
descriptorTag	8 bits;
descriptorLength	8 bits;
eventId	16 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;
}	

5.2. Implementação para o SBTVD

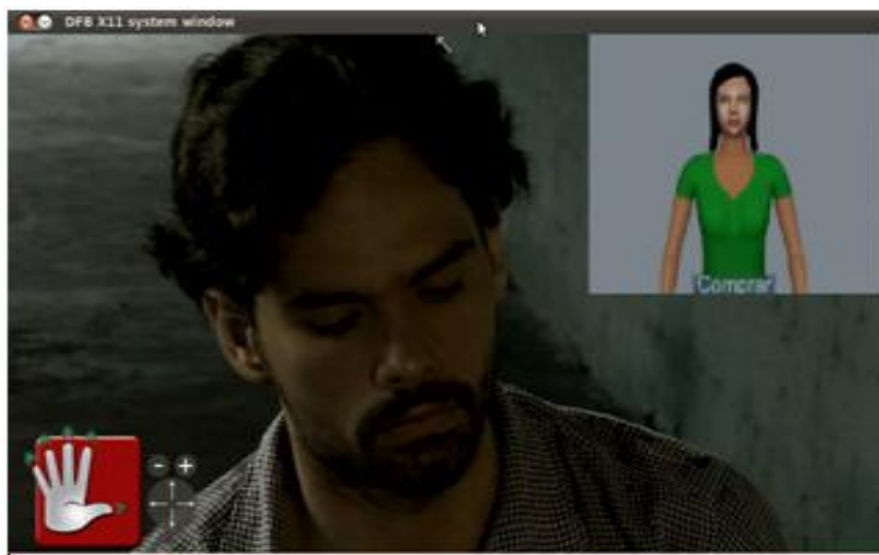
Inicialmente, foi desenvolvida utilizando a linguagem de programação C++, uma ferramenta responsável por codificar os sinais de LIBRAS denominado, Codificador de LIBRAS. Nessa implementação, um usuário digita textos em *glosa* (i.e., na estrutura gramatical de LIBRAS) que são codificados pela ferramenta,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

Para validar o fluxo gerado por essa ferramenta, do lado do receptor foi desenvolvida uma aplicação interativa Ginga-J [41] utilizando o Openginga, uma implementação de código aberto do middleware Ginga [41], disponível em [42]. Essa aplicação extrai as informações de LIBRAS do fluxo e, com o auxílio de um dicionário de LIBRAS (ver Capítulo 4), 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 [43].

A Figura 30 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.

Figura 30: App LIBRAS-TV



5.2.1. Representação de efeitos sonoros, emoções e identificação dos personagens

Para educação e alfabetização dos deficientes auditivos, são utilizados vários recursos, como por exemplo: desenho, produção coletiva, produção gestual e facial [44]. A Figura 31 exemplifica algumas imagens utilizadas no ensino aos surdos. Baseado nessa abordagem, o presente trabalho busca representar os efeitos sonoros produzidos nas cenas, através de desenhos, que irão servir como pistas visuais.

As pistas visuais são imagens que descrevem iconicamente os efeitos sonoros na cena atual, ou em segundo plano, conforme ilustrado na Tabela 7. 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.

Figura 31: Exemplo de imagens utilizadas na alfabetização de surdos, extraída de [44].

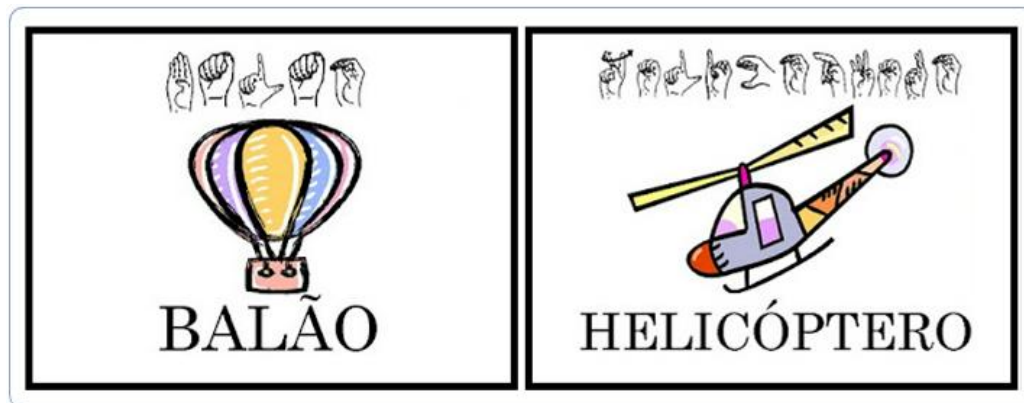
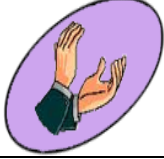
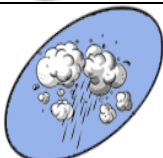


Tabela 7: Pistas visuais para sons do ambiente.

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

	Telefone
	Passos
	Sinos
	Gritos
	Chuva

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, podemos utilizar uma abordagem similar a de Ohene-Djan J. *et al.* [24], em que imagens com cores e números identificam quem está falando. Além disso, estas imagens também identificam 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 32.

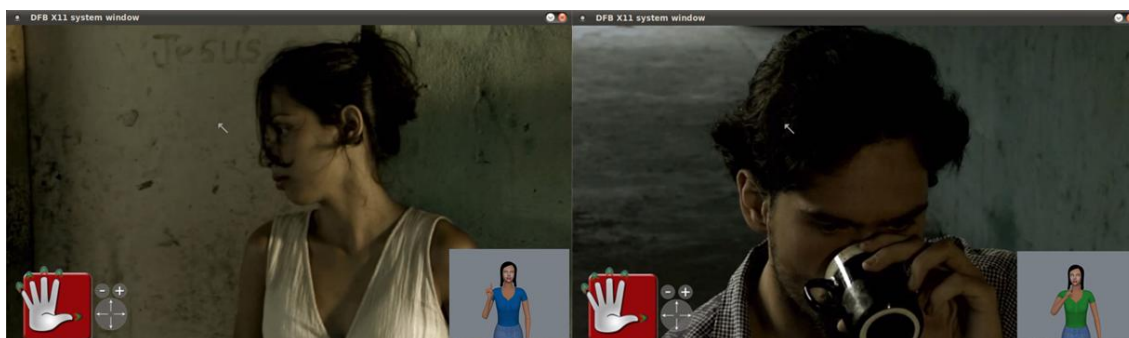
Figura 32: Imagens utilizadas para identificação do personagem.



Outras formas de identificar os personagens de um diálogo são através da inclinação do agente animado virtual e mudança de cor da roupa do agente animado

virtual a partir da troca do falante da cena. No primeiro caso, quando um personagem discursa, o avatar-3D se inclina em um sentido (para a esquerda, por exemplo) e quando outro personagem discursa o avatar-3D se inclina em outro sentido (para a direita, por exemplo). No segundo caso a camisa do intérprete muda de cor à medida que ocorre a mudança do personagem que está discursando, podemos observar um exemplo na Figura 33.

Figura 33: Identificação do personagem através da mudança de cor da roupa do intérprete.



As Figuras 34, 35 e 36 ilustram a representação de efeitos sonoros, de emoções e da identificação de personagens geradas pela aplicação interativa.

Figura 34: Demonstração do efeito sonoro (aplausos).

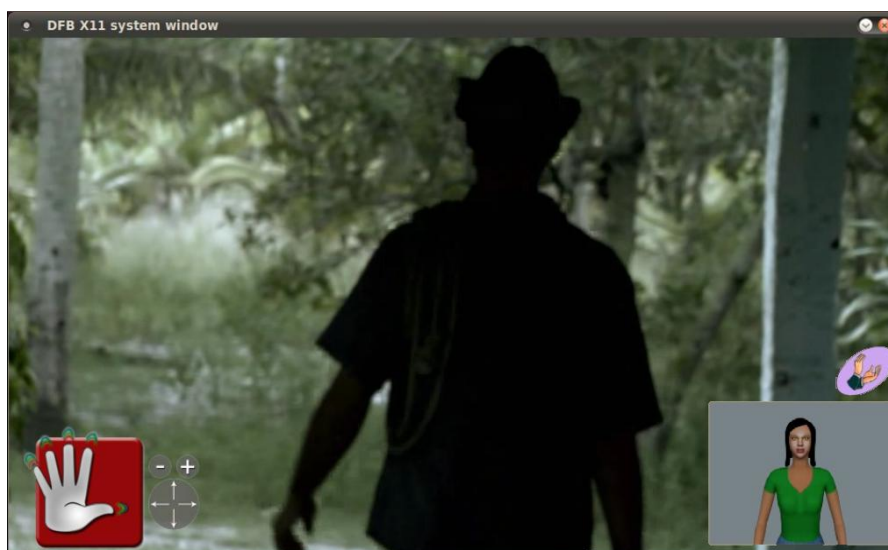


Figura 35: Demonstração de emoção sentida pelo personagem.

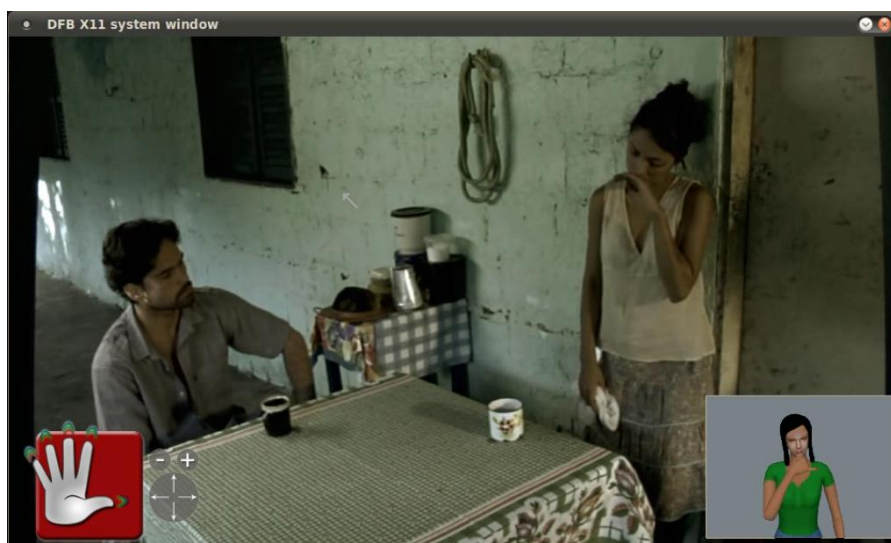
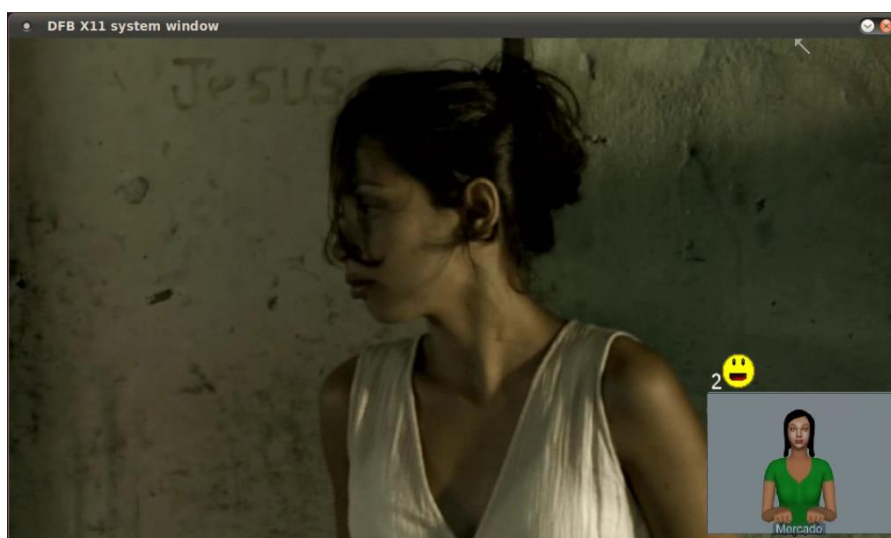


Figura 36: Identificação do personagem através de imagens.



5.3. Considerações Finais

Nesse capítulo foi apresentada a proposta de protocolo de codificação de janelas de LIBRAS para sistemas de TV digital. Nesta proposta, são transmitidos um conjunto de códigos, onde cada código está associado à representação visual de um sinal de LIBRAS, armazenada em um dicionário presente no receptor, possibilitando o suporte as diferenças regionais. Adicionalmente são sugeridas maneiras para representação de emoções, entonação de voz, humor e efeitos sonoros.

Também foram apresentadas as ferramentas utilizadas e os recursos necessários para a implementação do projeto e utilização do protocolo proposto neste trabalho. Foram exibidas algumas capturas de tela com o projeto em execução afim de demonstrar de forma mais concisa as informações descritas. No próximo capítulo são relatados testes que demonstram o funcionamento do sistema, comprovando que ele atende aos requisitos levantados nos capítulos anteriores e descritos neste capítulo sob aspectos objetivos e subjetivos.

6. RESULTADOS E DISCUSSÕES

Nesse capítulo, a solução proposta será avaliada através de um conjunto de testes computacionais e uma avaliação subjetiva conduzida com usuários surdos brasileiros. Na Seção 6.2, serão descritos alguns testes computacionais objetivos conduzidos para avaliar a largura de banda utilizada pelo protocolo proposto nesse trabalho. Na Seção 6.3 serão descritos alguns testes feitos com usuários surdos brasileiros conduzidos com o objetivo de avaliar a solução com relação ao nível de influência dos dialetos regionais, nível de compreensão dos efeitos sonoros e emoções, estratégias de identificação dos personagens, dentre outros.

6.1. Ambiente de Testes

Para realizar os primeiros testes com o protótipo, dois computadores (mini-PCs) com processador Intel Dual Core T3200 2GHz e 4GB de memória RAM, juntamente com alguns equipamentos de estação de TV, como, por exemplo, um Bombeador de Vídeo e um Multiplexador foram utilizados.

O primeiro computador foi utilizado para executar o módulo "Tradutor de LIBRAS" do protótipo e o segundo computador foi utilizado para executar a aplicação interativa LibrasTV. Uma versão do OpenSUSE foi instalada no segundo computador para tornar possível a execução da aplicação. O sistema operacional utilizado nos dois computadores foi o Linux Ubuntu 10.0.4, kernel 2.6.32.

Esses computadores e equipamentos de TV Digital foram integrados e organizados da seguinte forma:

- O Bombeador de Vídeo transmite via fluxo UDP alguns vídeos MPEG-2 TS de teste com *closed caption* embutido para o Tradutor de LIBRAS e para o Multiplexador;
- O "Gerador de LIBRAS" recebe o fluxo MPEG-2 TS do Bombeador de Vídeo e gera um fluxo de LIBRAS codificado, isto é, a sequência de glosas

e as informações de sincronização codificadas nas mensagens LDM e LCM e encapsuladas nos eventos DSM-CC;

- O Multiplexador recebe o fluxo MPEG-2 TS, o fluxo de dados com a aplicação do LibrasTV e o fluxo de LIBRAS codificado, remultiplexados em um único fluxo MPEG-2 TS e transmite esse fluxo via fluxo UDP para a máquina do Openginga (segundo computador);
- O Openginga recebe o fluxo MPEG-2 TS via fluxo UDP, extrai e exibe o vídeo do programa e executa a aplicação do LibrasTV que gera a trilha (janela) de LIBRAS a partir do fluxo de LIBRAS codificado.

6.2. Testes Computacionais Objetivos

Inicialmente, alguns testes computacionais foram realizados para avaliar a largura de banda utilizada pela solução. Nesse teste, dois vídeos foram transmitidos para o LibrasTV em sequência por 2 iterações e a taxa de bits (em Kbps) utilizada pelo fluxo de LIBRAS codificado (eventos DSM-CC com glosas encapsuladas) foi calculada. A taxa de bits (em Kbps) também foi calculada quando o vídeo com um intérprete de LIBRAS era transmitido para o usuário, em vez do fluxo de LIBRAS codificado. As Figuras 37, 38 e 39 mostram os resultados obtidos pelo LibrasTV e pela solução baseada em vídeos, respectivamente.

Nas Figuras 37 e 38 pode-se observar que a largura de banda de canal necessária para transmitir o fluxo de LIBRAS codificado (um valor médio de 5.64 Kbps e sempre inferior a 40 Kbps) foi bem inferior ao valor utilizado quando um fluxo de vídeo com intérprete de LIBRAS era utilizado (um valor médio de 3,64 Mbps)⁴, como podemos ilustrar na Figura 39. Um valor inferior de taxa de bits para o fluxo de LIBRAS codificado já era esperado, mas os testes indicaram que essa taxa é, em média, cerca de 600 vezes menor que o valor utilizado pelo fluxo de vídeo. O fluxo de vídeo com intérprete de LIBRAS utilizado no teste foi codificado utilizando um codificador de vídeo profissional MPEG-2 para uma resolução de 720 x 480. Outro fator importante a

⁴ Taxa obtida com o uso de um codificador MPEG-2, única opção permitida pela norma brasileira.

ser observado, é que a taxa do fluxo do vídeo foi alta mesmo durante os intervalos de silêncio. Os vídeos utilizados nessa parte do experimento são descritos na Tabela 8.

Tabela 8: Vídeos com *closed caption* utilizado no sistema.

Vídeos	Duração	Descrição
Video_1	26 sec.	Este vídeo é parte de um programa de notícias apresentado em 14 de Outubro 2008 na TV Globo.
Video_2	79 sec.	Este vídeo é parte de um programa de notícias apresentado em 31 de Janeiro 2007 na TV Globo.
Vídeo_3	65 sec.	Esse vídeo é parte de um filme produzido pela TV UFPB, desenvolvido com propósitos acadêmicos.

A taxa de bits absoluta usada para transmitir o fluxo de LIBRAS foi sempre inferior a 40 Kbps. A taxa de bits média foi de 5,37 Kbps para Video_1 e 5,57 Kbps para Video_2. Como podemos observar na Figura 37 e 38, a largura de banda utilizada pelo fluxo de LIBRAS é baixa, assim, pode ser possível a aplicação do processo proposto em outras plataformas de rede, tais como a Web. Além disso, esta banda foi significativamente menor do que a largura de banda utilizada se a opção for transmitir a janela de LIBRAS como um vídeo em vez de fluxo de LIBRAS codificado, como podemos observar nos trabalhos de Dimitris Agrafiotis et al. [30] e Richard P. Schumeyer et al. [29], em que a taxa variava entre 40 e 50 kbps. A qualidade foi questionada em avaliações subjetivas.

Figura 37: Largura de banda (em kbps) utilizada pelo fluxo de LIBRAS (Video_1).

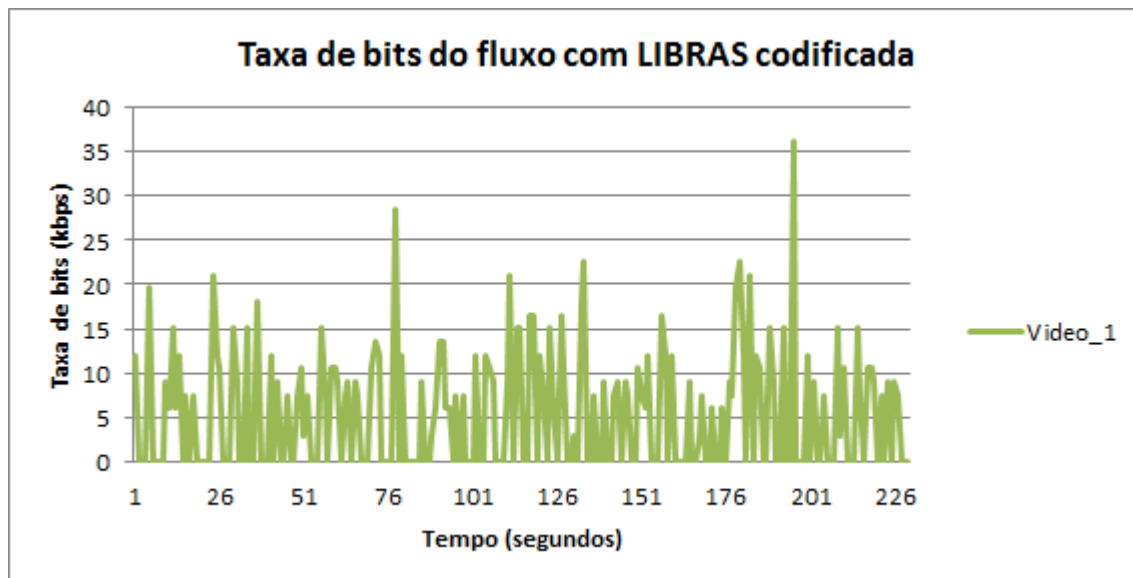


Figura 38: Largura de banda (em kbps) utilizada pelo fluxo de LIBRAS (Video_2).

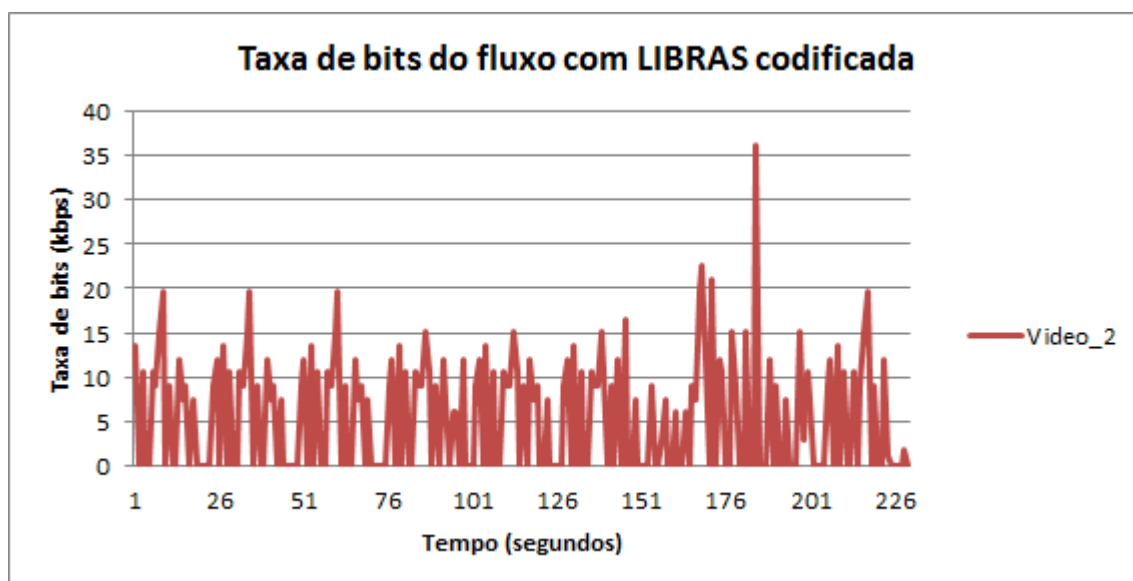
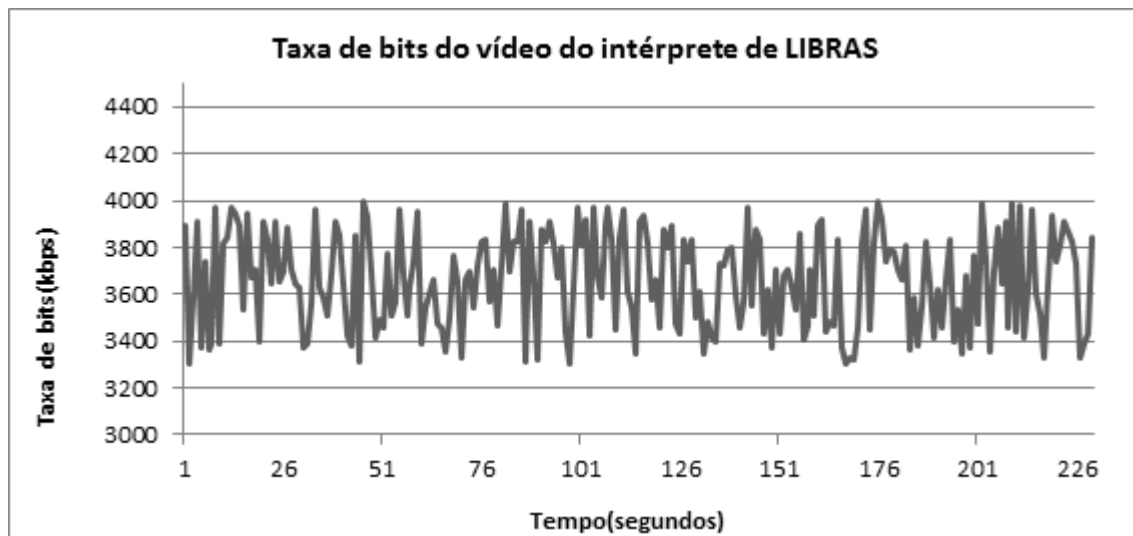


Figura 39: Largura de banda (em kpbs) utilizada pelo fluxo de vídeo com intérprete de LIBRAS (vídeo_3).



Para avaliar se as diferenças de desempenho entre a transmissão do *closed caption* e a transmissão da LIBRAS codificada com o protocolo, são estatisticamente representativas, foi aplicado um teste para a média de duas amostras independentes. O teste estatístico aplicado foi o Teste-t de Student [45], um dos testes paramétricos mais utilizados na literatura científica para comparar dois tratamentos. Os resultados apresentados mostram que o protocolo de transmissão de LIBRAS descrito nesse trabalho ocupa uma banda de transmissão média menor que a técnica de CC e essa diferença é estatisticamente significativa ($p\text{-valor} < 0,05$), como podemos observar na Tabela 9.

Tabela 9: Resultados do teste de validação de hipótese para compreensão dos conteúdos

Amostra	Taxa Média	Variância	Grau de liberdade	p-valor (bilateral)
CC	6.9	44.94	451	0.0117
Protocolo	5.3	44.06		

Também foram feitos alguns testes com o objetivo de medir os atrasos de cada módulo do projeto LibrasTV. De acordo com a Tabela 10, o prazo médio para executar todos os componentes do LibrasTV (ou seja, a soma de extração do *closed caption*, tradução automática, codificação, decodificação e exibição) era de menos de 43

milissegundos. O atraso máximo obtido (considerando-se o atraso máximo de cada módulo) foi 142,26 milissegundos, enquanto que o atraso mínimo foi de 20,509 milissegundos. Considerando que o teste foi realizado com um vocabulário aberto e representativo e num cenário real, as janelas de LIBRAS podem ser geradas em tempo real e com um tempo médio menor do que o tempo necessário para uma tradução humana. Além disso, este tempo de atraso médio é menor do que o tempo utilizado em outros trabalhos, tais como solução proposta por San-Segundo et al. [46][47][48], que relataram um atraso médio em torno de 8 segundos por frase.

Tabela 10: Medida das médias de atraso de cada módulo do LibrasTV.

	Média (ms)	Desvio Padrão (ms)	Maximo (ms)	Mínimo (ms)
Extração CC	0,024	0,022	0,554	0,017
Tradução	0,975	2,957	80,126	0,220
Codificação	0,215	0,089	1,061	0,072
Decodificação	0,170	0,143	0,519	0,020
Exibição	42,445	8,747	59,998	20,000
Total	43,805	-	142,21	20,509

6.3. Testes com os Usuários

Após implementar a solução, alguns testes com usuários surdos e intérpretes de LIBRAS foram aplicados para avaliar a viabilidade e adequabilidade da solução. Os testes foram realizados em dois conjuntos de testes. No primeiro conjunto, os testes com usuários foram realizados com o objetivo de avaliar o grau de relevância do sistema, a regionalização, a exibição de informações não literais, dentre outros. No segundo grupo de testes, os usuários avaliaram a qualidade dos conteúdos gerados considerando um conjunto maior de conteúdos multimídia, e também examinaram algumas estratégias de identificação do falante (personagem) da cena.

6.3.1. Instrumentação

Conforme sugerido por Eronen [2006] [49], entrevistas usando métodos de observação são importantes em estudos exploratórios, pois fornecem dados sobre o comportamento do visualizador e possíveis explicações para esse comportamento. Para estudar as reações do espectador a um novo conceito, LIBRAS com legendas reforçadas, foi realizada uma primeira etapa de testes com 12 surdos e 2 intérpretes de LIBRAS de João Pessoa/Paraíba. O grupo foi formado por homens e mulheres, entre 16 e 36 anos, e com variados grau de instrução. Os usuários foram convidados a assistir cinco vídeos (ver Tabela 11) e responder a um questionário com o objetivo de avaliar os seguintes aspectos: diferenças regionais, uso correto da LIBRAS, representação da emoção e representação de efeitos sonoros. Um intérprete de LIBRAS estava presente durante os testes para auxiliar os surdos no preenchimento do questionário de avaliação.

Tabela 11: Conteúdos multimídias usados na primeira bateria de testes.

Vídeo	Objetivo	Duração
Trecho de um filme produzido pela TV UFPB, desenvolvido com fins acadêmicos, que representa um diálogo entre dois personagens.	Avaliar o grau de relevância do sistema proposto	65 seg.
Trecho de um filme desenvolvido pela TV UFPB, produzido em duas versões: (1) com a janela de LIBRAS com sinais de João Pessoa e (2) com a janela de LIBRAS com sinais de São Paulo.	Avaliar o nível de entendimento dos surdos com sinais de sua região de origem	65 seg.
Trecho de um filme desenvolvido pela TV UFPB, com identificação de personagem, produzido em duas versões	Avaliar o entendimento dos surdos sobre quem está falando na cena	65 seg.
Cenas representando emoções	Avaliar o nível de entendimento dos surdos sobre as emoções sentidas pelos personagens	20 seg.
Cenas demonstrando efeitos sonoros	Avaliar o nível de entendimento dos surdos sobre as pistas visuais que representam efeitos sonoros	20 seg.

A segunda etapa de testes foi realizada com 20 surdos de LIBRAS de João Pessoa/Paraíba. O grupo foi formado por homens e mulheres, entre 13 e 56 anos, também com variados grau de instrução. Os usuários assistiram a 5 vídeos e responderam a um questionário com o objetivo de avaliar os seguintes aspectos:

qualidade dos conteúdos gerados e grau de entendimento sobre quem estava falando na cena.

De acordo com uma pesquisa realizada pela Associação Brasileira de Rádio e Televisão (ABERT) e pela Fundação Getúlio Vargas (FGV) [50], a TV Brasileira possui uma programação bastante diversificada envolvendo conteúdos de diversos gêneros que são distribuídos da seguinte forma: filmes, seriados e novelas (35,3% da programação), programas jornalísticos (20,3% da programação), programas infantis (14,1% da programação), programas de variedade (12,5% da programação), programas esportivos (5,0% da programação), programas educacionais (2,6% da programação), programas de comédia (2,5% da programação), programas religiosos (2,0% da programação), reality shows (1,4% da programação), etc. Com base nessa pesquisa, os conteúdos multimídia selecionados como objetos deste experimento são quatro conteúdos multimídia que pertencem a diversos gêneros, como por exemplo, trechos de vídeos de noticiários, de filmes, de programas esportivos, de programas de auditório, dentre outros, priorizando dessa forma os gêneros mais representativos da TV Brasileira. Na Tabela 12, são apresentados os conteúdos multimídia utilizados nos testes.

Além disso, foram escolhidos conteúdos curtos para evitar que os testes sejam muito longos e os usuários se entediem durante a avaliação, influenciando os resultados. Para avaliar o nível de entendimento acerca de quem estava falando na cena, foi utilizado um trecho de um filme desenvolvido pela TV UFPB, produzido em duas versões: (1) com identificação através da inclinação do avatar-3D e (2) com identificação através da mudança da roupa do avatar-3D.

Tabela 12: Conteúdos multimídias usados na segunda bateria de testes.

Vídeo	Objetivo	Duração
Trecho de um filme produzido pela TV UFPB, desenvolvido com fins acadêmicos, que representa um diálogo entre dois personagens.	Avaliar a qualidade dos conteúdos gerados e o grau de relevância do sistema	65 seg.
Trecho de programa jornalístico veiculado pela TV Globo	Avaliar a qualidade dos conteúdos gerados e o grau de relevância do sistema	26 seg.
Trecho do programa de auditorio Hoje em Dia veiculado pela Rede TV	Avaliar a qualidade dos conteúdos gerados e o grau de relevância do sistema	98 seg.
Filme produzido pela Blender Foundation	Avaliar a qualidade dos conteúdos gerados e o grau de relevância do sistema	896 seg.

Trecho de um filme desenvolvido pela TV UFPB, com identificação de personagem, produzido em duas versões	Avaliar o entendimento dos surdos sobre quem está falando na cena	65 seg.
--	---	---------

Os dados dessa avaliação também foram coletados através de questionários. Os questionários usados nas avaliações são compostos de duas partes e são apresentados no Anexo A. Na primeira parte do questionário, o usuário fornece informações pessoais como sexo, idade, tipo de surdez, grau de escolaridade, dentre outros e na segunda parte, ele avalia cada um dos conteúdos apresentados respondendo algumas questões sobre os vídeos exibidos.

6.3.2. Primeiro grupo de testes

Essa parte do experimento foi conduzida no Shopping Tambiá e os sujeitos foram um conjunto de 12 surdos brasileiros e 2 intérpretes de LIBRAS. Neste teste, os usuários (todos naturais de João Pessoa) foram divididos em dois grupos de mesmo tamanho de forma aleatória. Os usuários foram convidados a assistir um trecho de um filme representando um diálogo entre duas pessoas e posteriormente submetidos a um questionário contendo 12 questões com 4 alternativas cada, com o objetivo de avaliar diferentes aspectos da solução proposta. Os usuários possuíam idade entre 16 e 36 anos, e idade média de 25 anos, sendo 5 do sexo feminino e 9 do sexo masculino.

A seguir descrevemos os passos que foram seguidos para realização dos testes com os deficientes auditivos:

1. Os usuários responderam questões pessoais, tais como idade, sexo e nível de instrução em LIBRAS e língua portuguesa;
2. Posteriormente os usuários foram convidados a avaliar o seu grau de satisfação com a solução e fazer uma auto-avaliação do entendimento do conteúdo. Os usuários responderam questões para avaliação da qualidade e naturalidade dos movimentos produzidos, velocidade de apresentação, nível de entendimento e os principais problemas da solução proposta.

Um dos aspectos avaliados foi o nível de influência da regionalização. Para isso dividimos o conjunto de surdos em dois grupos de mesmo tamanho. Um grupo assistiu e avaliou um vídeo com os sinais da região de São Paulo e o outro assistiu e avaliou um vídeo com os sinais da região de João Pessoa. Os vídeos utilizados nessa parte do experimento estão descritos na Tabela 13.

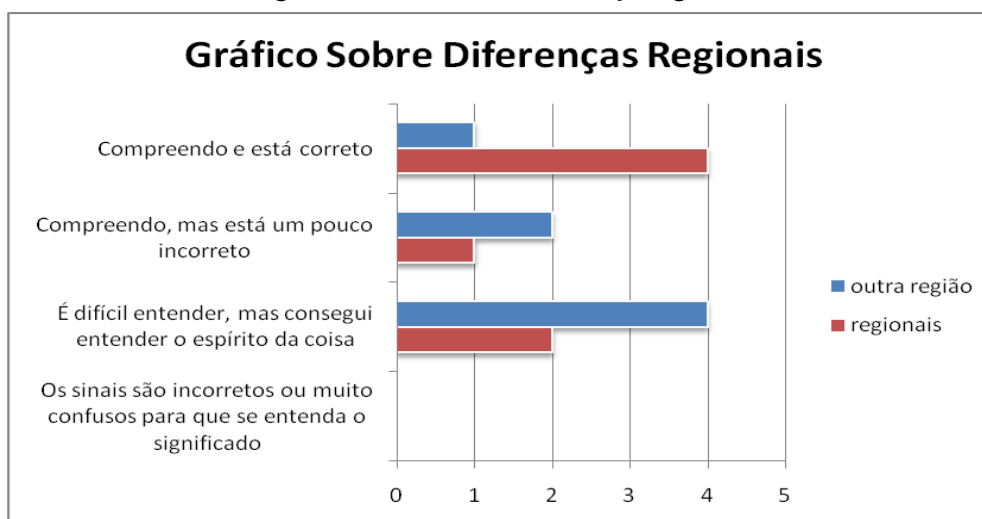
Tabela 13: Vídeos utilizados para avaliação das diferenças regionais.

Videos	Duração	Descrição
Video_1	65 sec.	Trecho de um filme produzido pela TV UFPB, desenvolvido com fins acadêmicos, que representa um diálogo entre dois personagens. Os sinais realizados pelo avatar-3D são referentes a região de São Paulo.
Video_2	65 sec.	Trecho de um filme produzido pela TV UFPB, desenvolvido com fins acadêmicos, que representa um diálogo entre dois personagens. Os sinais realizados pelo avatar-3D são referentes a região de João Pessoa.

Na Figura 40, são apresentados os resultados obtidos na avaliação das diferenças regionais, no qual o eixo ‘X’ do gráfico representa o número de surdos que optaram pelas alternativas existentes no eixo ‘Y’.

Conforme pode ser observado na Figura 40, observou-se que, de acordo com a avaliação dos usuários, as diferenças regionais influenciam na compreensão do conteúdo representado. Os usuários que assistiram ao vídeo com os sinais de João Pessoa apontaram que compreenderam melhor o contexto do filme que os usuários que assistiram ao vídeo com os sinais de São Paulo.

Figura 40: Gráfico sobre as diferenças regionais.

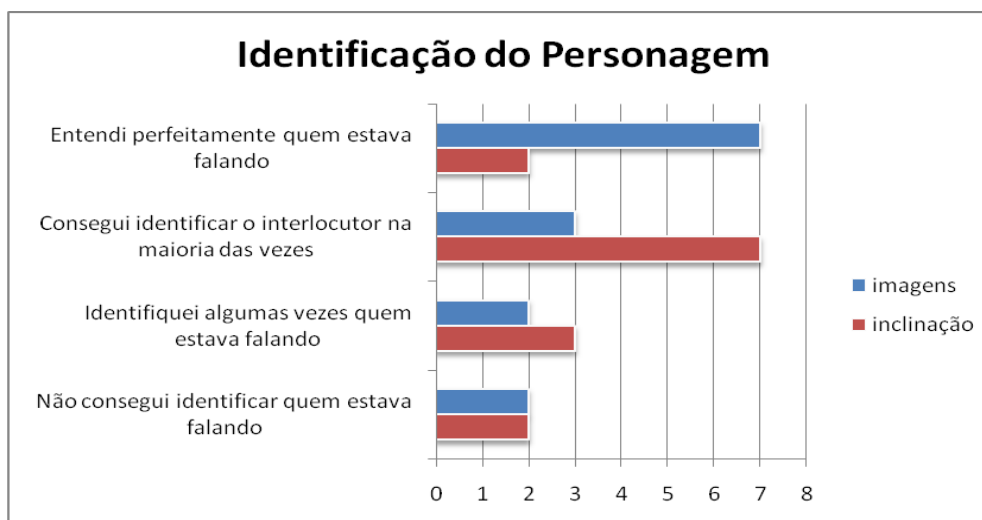


Na Figura 41, são apresentados os resultados obtidos com relação à identificação dos personagens. Nesse teste, os usuários foram divididos em dois grupos: o primeiro grupo avaliou um vídeo com identificação do personagem através de imagens (*emotions*) acompanhadas do alfabeto *alfa-numérico* e o segundo grupo avaliou um vídeo com identificação do personagem através da inclinação do avatar-3D. Conforme pode ser observado na Figura 41, um número mais elevado de usuários indicou um maior grau de compreensão com a identificação do personagem através do uso de imagens. Os vídeos utilizados nessa parte do experimento são descritos na Tabela 14.

Tabela 14: Vídeos utilizados para avaliar a identificação do personagem falante.

Vídeos	Duração	Descrição
Video_1	65 sec.	Trecho de um filme produzido pela TV UFPB, desenvolvido com fins acadêmicos, que representa um diálogo entre dois personagens. A identificação do personagem é demonstrada através da utilização de imagens (<i>emotions</i>) com cores diferentes e acompanhadas do alfabeto <i>alfa-numérico</i> .
Video_2	65 sec.	Trecho de um filme produzido pela TV UFPB, desenvolvido com fins acadêmicos, que representa um diálogo entre dois personagens. A identificação do personagem é demonstrada através da inclinação do avatar-3D.

Figura 41: Gráfico sobre a identificação do personagem.



As pistas visuais não tiveram um resultado satisfatório, obtendo uma média de aprovação de 43,75%. A maioria dos usuários indicaram que não entenderam ou confundiram o entendimento sobre o que a frase estava querendo demonstrar. As emoções representadas pelo avatar-3D através de expressões faciais tiveram um resultado positivo, obtendo uma média de aprovação de 81,5%.

Um dos problemas detectados nessa primeira etapa de testes foi a naturalidade do sinal executado pelo avatar-3D. De acordo com os usuários, os sinais possuíam pouca flexibilidade e os movimentos não eram naturais. Para superar esse problema, a ferramenta de captura de movimentos “Microsoft Kinect”⁵ foi integrada com as ferramentas de modelagem do avatar-3D. Com isso, foi possível observar que a sinalização ficou mais natural e flexível. No segundo conjunto de testes, todos os sinais utilizados no primeiro conjunto de testes, foram novamente gerados com base na captura de movimentos usando o Kinect integrado com o Blender [43] realizados por uma deficiente auditiva.

O protocolo proposto oferece duas opções de transmissão: (1) a transmissão do vídeo de um intérprete humano ou (2) a janela de língua de sinais realizada pelo avatar-3D. Com isso fica claro que a intenção do projeto LibrasTV, assim como o protocolo tema deste trabalho, não é a substituição do intérprete de língua de sinais pelo avatar-3D e sim a utilização do agente animado virtual na ausência da janela de língua de sinais com a sinalização realizada por um intérprete humano.

⁵ www.xbox.com/kinect

Com o intuito de avaliar o nível de aceitação dos surdos com a exibição da LIBRAS (língua de sinais foco desse trabalho) realizada pelo agente virtual animado 3D, os usuários avaliaram o grau de relevância da solução proposta e a capacidade de uso do sistema na ausência de intérprete humano. A Figura 42 e 43 demonstram os resultados obtidos.

Figura 42: Gráfico que demonstra a opinião dos surdos sobre o grau de relevância do sistema proposto.

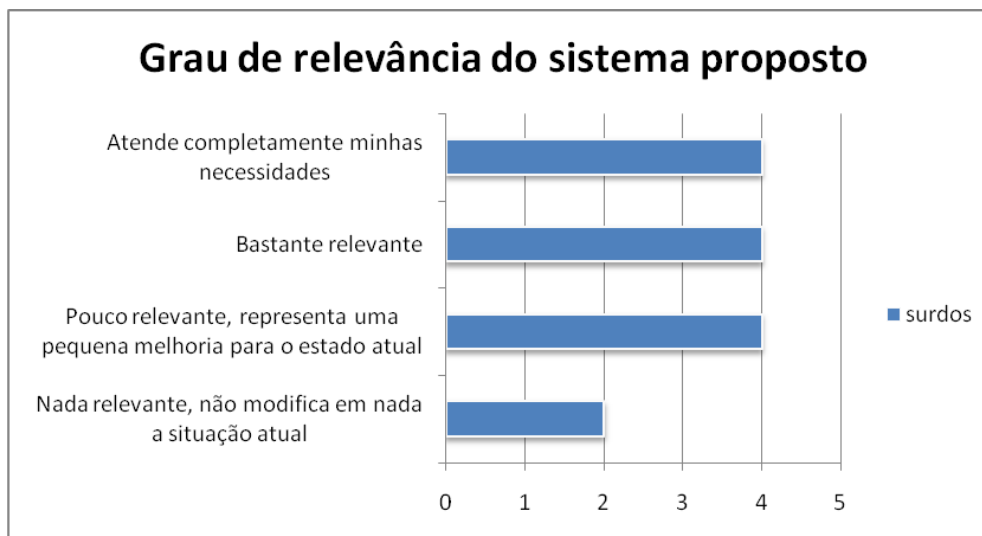
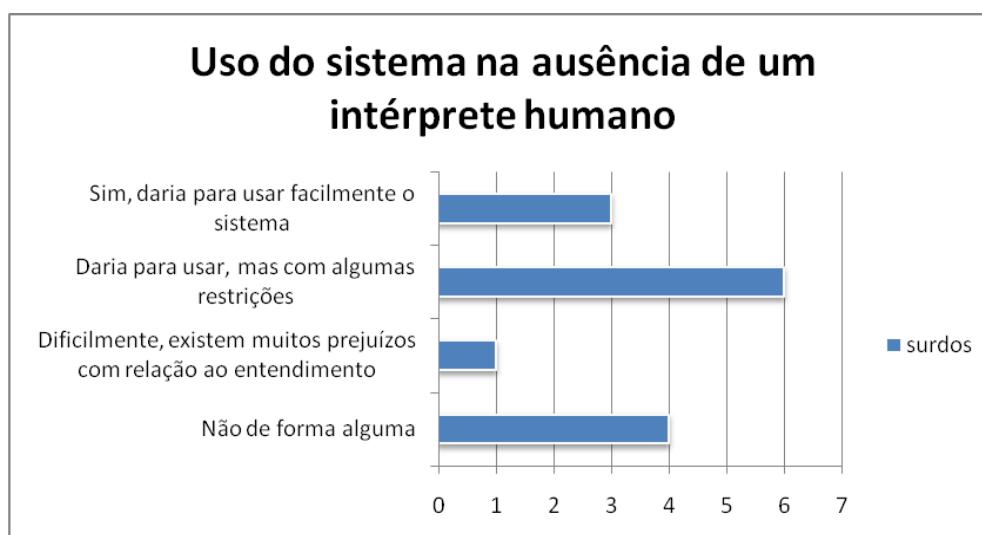


Figura 43: Gráfico que demonstra a aprovação dos surdos para o uso do sistema na ausência de um intérprete humano.



6.3.3. Segundo grupo de testes

O segundo conjunto de testes com usuários foi realizado na Fundação Centro Integrado de Apoio a Pessoa com Deficiência (FUNAD). Os usuários foram convidados a assistir quatro vídeos (demonstrados na subseção 6.1.1), de tipos de conteúdos televisivos diferentes, sem que estes recebessem qualquer tipo de compensação. Sendo assim, 20 (vinte) surdos de diferentes níveis de instrução e idades variadas foram divididos em dois grupos e ambos os grupos assistiram aos mesmos vídeos. Enquanto um destes grupos pôde ver a apresentação do *closed caption*, o outro pôde assistir o conteúdo em LIBRAS sendo executado por um avatar-3D, conteúdo gerado de acordo com a solução proposta neste trabalho.

O grupo de usuários era composto por onze mulheres e nove homens na faixa etária de 13 a 56 anos, e com uma média de idade de 28,6 anos. O grau de escolaridade e o nível de conhecimento em Português e LIBRAS dos usuários também foi observado e é apresentado na Figura 44 e na Tabela 15, respectivamente.

Figura 44: Grau de escolaridade dos usuários que participaram do teste.

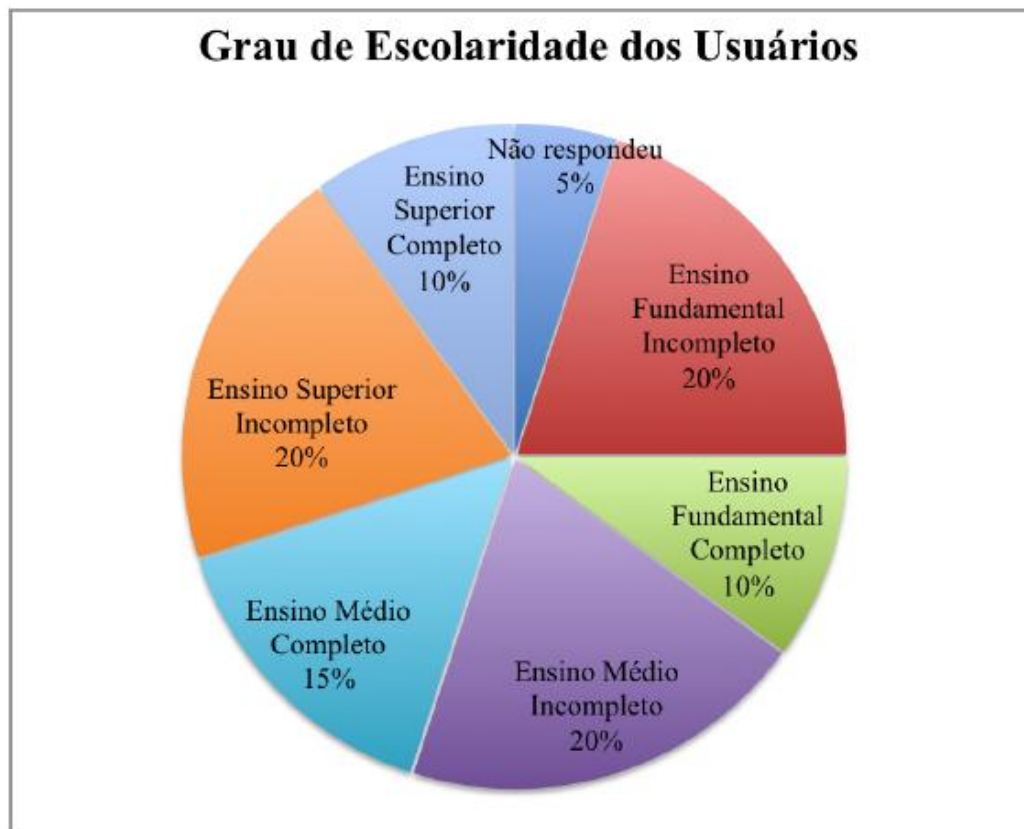


Tabela 15: Nível de conhecimento em português e em LIBRAS dos usuários.

Nível de conhecimento	Valor Médio	Desvio Padrão
Português (1 a 6)	4,00	1,80
LIBRAS (1 a 6)	5,67	1,54

A dinâmica para aplicação dos testes ocorreu da seguinte forma:

1. Os usuários responderam questões pessoais, tais como idade, sexo e nível de instrução em LIBRAS e língua portuguesa;
2. Após assistir cada vídeo os usuários foram instruídos a responder perguntas sobre o conteúdo do vídeo exibido. É importante destacar que estas perguntas tinham por objetivo gerar uma medida objetiva sobre o nível de compreensão dos usuários a respeito do vídeo exibido. Esta porção do questionário possuía, ao total, 16 questões sobre os quatro vídeos apresentados;
3. Por fim os usuários foram convidados a avaliar o seu grau de satisfação com a solução e fazer uma auto-avaliação do entendimento do conteúdo.

Os usuários que foram submetidos ao teste com o avatar-3D responderam também questões para avaliação da qualidade e naturalidade dos movimentos produzidos. Algumas fotos do experimento são apresentadas na Figura 45.

Figura 45: Fotos do experimento com os usuários.



Durante a execução dos testes e preenchimento dos questionários, os usuários tiveram o auxílio de um intérprete profissional de LIBRAS, funcionário da FUNAD, como forma de entender melhor as perguntas presentes na avaliação. As Tabelas 16 e 17 apresentam os níveis de compreensão médios dos usuários para cada um dos vídeos testados. A Tabela 16 apresenta os resultados para os usuários que assistiram os vídeos com *closed caption*, enquanto a Tabela 17 mostra os níveis de compreensão para usuários testados com o avatar-3D.

Tabela 16: Níveis de compreensão para vídeos com *closed caption*.

Conteúdo	Percentual de acertos	Número de acertos médio	Desvio-padrão
Vídeo_1	27%	0.8	0.75
Vídeo_2	15%	0.6	1.02
Vídeo_3	26%	1.3	1.27
Vídeo_4	35%	1.4	0.91
Total	26%	4.1	3.01

Tabela 17: Níveis de compreensão para vídeos com janela de LIBRAS com avatar-3D.

Conteúdo	Percentual de acertos	Número de acertos médio	Desvio-padrão
Vídeo_1	90%	2.7	0.45
Vídeo_2	62%	2.5	0.67
Vídeo_3	84%	4.2	0.74
Vídeo_4	82%	3.3	0.78
Total	79%	12.7	1.42

As Tabelas mostram o número de questões respondidas de forma correta como também o percentual médio de acertos dos usuários. Cada uma das perguntas possuía quatro alternativas, com apenas uma resposta correta, e uma alternativa que poderia ser marcada caso os usuários não soubessem a resposta para a pergunta. Esta alternativa foi acrescida para evitar que os usuários fizessem escolhas aleatórias quando não soubessem a resposta.

Como pode ser observado nas Tabelas 16 e 17 houve um significativo aumento no percentual de acerto das questões quando os usuários assistiram ao vídeo com conteúdo em LIBRAS gerado usando o avatar-3D. O menor desvio-padrão encontrado para o resultado da avaliação com o avatar-3D pode ser causado por uma uniformidade maior do conhecimento de LIBRAS do que no conhecimento da língua portuguesa. Isto acontece porque muitos surdos são analfabetos, e 97% deles não completam o ensino médio. Isto faz com que sua capacidade de leitura seja muito prejudicada e varie muito entre surdos com maior e menor nível de instrução em língua portuguesa.

Para verificar se as diferenças de desempenho entre as soluções utilizando *closed caption* e a LIBRAS codificada são representativas um teste Mann-Whitney foi aplicado. Com base na Tabela 18 concluímos que a diferença no entendimento por parte dos surdos favorável à solução proposta é estatisticamente relevante ($p\text{-valor} < 0.05$). Considerando-se que vários tipos diferentes de conteúdo foram testados, este resultado indica que o uso de janela de LIBRAS com avatares representa uma melhora em relação ao cenário atual em que se encontram os surdos brasileiros. O teste foi feito utilizando o software SPSS, versão 17.0.

Tabela 18: Resultados do teste de validação de hipótese para compreensão dos conteúdos multimídias.

Tipo de amostra	Média de acertos	Mediana	Desvio-padrão	p-valor (bilateral)
<i>Closed caption</i>	4.1	3.5	3.01	0.0005
Avatar-3D	12.7	12	1.42	

Os usuários foram requisitados a fazer uma auto-avaliação (com notas entre 1 e 6) sobre seu conhecimento de língua portuguesa e LIBRAS. Essa escala (de 1 a 6) foi escolhida porque, de acordo com Morrissey (2008), ela estimula os usuários a avaliarem positivamente ou negativamente os fatores, evitando avaliações neutras. Apesar da avaliação ter sido feita pelos próprios usuários, e isto poder gerar uma tendência de avaliação positiva, também aplicamos o teste de Mann-Whitney para avaliar a diferença no nível declarado do conhecimento de LIBRAS e de língua portuguesa entre os dois grupos testados. Este é um teste de hipótese estatística não-paramétrica, utilizado quando há dois grupos independentes de amostras (já que dois grupos testarão cada uma das soluções), em que o valor das variáveis dependentes tem caráter ordinal. A Tabela 19 mostra os resultados para a avaliação do conhecimento em LIBRAS e a Tabela 20 mostra os valores para conhecimento de língua portuguesa. As Tabelas mostram que não existem diferenças estatisticamente significativas ($p\text{-valor} > 0.05$) entre o conhecimento declarado por usuários dos dois grupos.

Tabela 19: Auto-avaliação dos usuários para conhecimento em LIBRAS.

Grupo avaliado	Média	Mediana	Desvio-padrão	p-valor (bilateral)
<i>Closed caption</i>	5.0	6.0	2.0	0.8798
Avatar-3D	5.8	6.0	0.4	

Tabela 20: Auto-avaliação dos usuários para conhecimento em língua portuguesa.

Grupo avaliado	Média	Mediana	Desvio-padrão	p-valor (bilateral)
<i>Closed caption</i>	3.9	4.0	1.91	0.2265
Avatar-3D	4.7	5.0	1.45	

Quando os usuários surdos foram questionados sobre a naturalidade do agente virtual animado 3D, o valor médio obtido foi de 4,4 como podemos ver na Tabela 21. Como mencionado em trabalhos anteriores [51][52], abordagens utilizando avatares não são bem vistas por uma boa parte dos surdos que tem como preferência a sinalização humana. Uma das razões para esta preferência, de acordo com Kipp et al. [52], é a dificuldade das abordagens de sinalização virtual para representação de emoções e movimentos com menos rigidez.

Tabela 21: Avaliação dos usuários para janela de LIBRAS com avatar-3D.

Critério avaliado	Média	Desvio-padrão
Naturalidade	4.4	1.39
Movimento das mãos	4.8	1.40
Expressões faciais	4.6	1.41

O critério de “naturalidade” avalia se os movimentos produzidos pelo avatar-3D se assemelham aos produzidos por um intérprete humano ou se não suficientemente fluidos e flexíveis. O critério de “movimento das mãos” tem o objetivo de saber se os usuários consideram que os movimentos apresentados pelo avatar-3D estão corretos ou não, enquanto que o critério “expressões faciais” mede a opinião dos usuários a respeito da correte e intensidade das expressões faciais presentes no conteúdo apresentado como também das emoções expressadas pelos personagens.

Os valores encontrados mostram que todos os critérios tiveram avaliação positiva, porém, devem ser investidos mais esforços com objetivo de tornar os sinais os mais naturais possíveis.

Para avaliar o grau de identificação dos personagens falantes de uma cena, o conjunto de surdos também foi dividido em dois grupos. Cada grupo assistiu a um vídeo que demonstrava o diálogo entre um homem e uma mulher. Posteriormente os usuários foram convidados a responder 4 questões com perguntas referente as falas dos personagens. Com isso, nós pudemos avaliar se de fato os usuários conseguiram identificar o falante de cada frase.

Um grupo assistiu e avaliou um trecho de um filme com a identificação do personagem através da inclinação do avatar-3D e outro grupo assistiu e avaliou o mesmo trecho de filme, porém, com a identificação do personagem realizada através da

mudança de cor da roupa do avatar-3D. Os vídeos utilizados nessa parte do experimento são apresentados na Tabela 22.

Tabela 22: Vídeos utilizados para identificação do personagem.

Videos	Duração	Descrição
Video_1	65 sec.	Trecho de um filme produzido pela TV UFPB, desenvolvido com fins acadêmicos, que representa um diálogo entre dois personagens. A identificação do personagem é demonstrada através da inclinação do avatar-3D.
Video_2	65 sec.	Trecho de um filme produzido pela TV UFPB, desenvolvido com fins acadêmicos, que representa um diálogo entre dois personagens. A identificação do personagem é demonstrada através da mudança de cor da roupa do avatar-3D

A Tabela 23 mostra os resultados da avaliação da identificação do personagem falante da cena. Esta Tabela mostra os valores para número de acertos médios e uma análise estatística dos resultados. Para essa etapa dos testes também foi aplicado Mann-Whitney para avaliar as duas formas de identificação sugeridas. A forma natural de identificação do personagem, utilizadas por intérpretes humanos, é através da inclinação do corpo, porém, conforme pode ser visto na Tabela 23, observou-se que não houve uma diferença estatisticamente significativa entre as duas abordagens utilizadas (p-valor > 0.05).

Tabela 23: Avaliação dos usuários sobre formas de identificação do personagem.

Grupo avaliado	Média	Mediana	Desvio-padrão	p-valor (bilateral)
Inclinação	1.9	2.0	1.2	0,3538
Cor	1.3	1.0	0.8	

6.4. Considerações Finais

Este capítulo apresentou um conjunto de testes objetivos e testes com usuários realizados para avaliar a solução proposta. Os testes ajudaram a avaliar alguns aspectos da solução como as estratégias de identificação do personagem, o nível de compreensão

dos conteúdos, dentre outros. A partir desses testes, foi possível observar que o protocolo proposto utiliza pouca largura de banda do canal de transmissão, é capaz de representar emoções e implementar estratégias de identificação do personagem em janelas de LIBRAS, dentre outros.

O protocolo proposto nesse trabalho permite o suporte a especificidades regionais da LIBRAS, uma vez que diferentes dicionários podem ser utilizados na solução dependendo da região onde ela é aplicada, contemplando deficientes auditivos com variados níveis de instrução em LIBRAS. Os dialetos regionais podem influenciar de maneira significativa a comunicação entre os deficientes auditivos. Comumente, os mais afetados pelo regionalismo, costumam ser crianças ou adultos que ainda estão em processo de alfabetização. Com isso, temos como principal contribuição desse trabalho, o apoio às diferenças regionais. Com essa solução, todos os deficientes auditivos podem ter o acesso a informação através dos conteúdos televisivos, e a comunicação e o entendimento podem fluir de maneira natural.

7. CONCLUSÃO E CONSIDERAÇÕES FINAIS

Este capítulo apresenta à avaliação geral do protocolo, sua utilização em sistemas de TV e cinemas digitais, as contribuições feitas, os resultados alcançados, trabalhos futuros e sugestões para melhorá-lo.

7.1. Resultados Alcançados

A solução proposta permite transmitir os sinais a baixas taxas de bits, além de permitir que informações relacionadas a efeitos sonoros, emoções, e a identificação do personagem que está discursando também sejam transmitidas, mesclando as funcionalidades presentes nos trabalhos vistos anteriormente, e adicionalmente, permitindo que as diferenças regionais presentes na LIBRAS sejam respeitadas. Alguns testes foram aplicados com surdos para validar a solução proposta. Nesses testes, foi verificada uma tendência de 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 tendem a não atender às necessidades dos usuários surdos.

7.2. Trabalhos Futuros e Sugestões

Dentre as propostas de trabalho futuro, está a de validar a solução em outros cenários, como em cinemas digitais e teatros, e desenvolver outras estratégias para identificação dos personagens e representação de efeitos sonoros para que possa ser alcançado o nível mais alto de compreensão.

Os efeitos sonoros produzidos em cena, ou em segundo plano, podem ser relevantes para o entendimento do contexto da programação que esta sendo exibida, porém, atualmente, essa informação muitas vezes é descartada, passando despercebida para os deficientes auditivos. Possíveis evoluções no mecanismo de exibição das

informações não literais (como por exemplo, efeitos sonoros) serão necessárias para que possamos oferecer um alto grau de qualidade e compreensão ao público deficiente auditivo.

Outro ponto relevante que deve ser investigado com mais ênfase é a geração automática da janela de LIBRAS a partir de uma entrada de áudio. Todos os testes realizados nessa proposta tiveram como ponto de partida legendas em Português Brasileiro, porém, temos como trabalho futuro o desenvolvimento ou a integração de reconhecedores de voz que irão captar o áudio, transformar em texto e passar por todo processo de geração automática da janela de LIBRAS realizado pelo LibrasTV.

7.3. Contribuições

As principais contribuições desse trabalho foram:

- Criação de um protocolo que permite a transmissão da língua de sinais para sistemas de televisão digital, cinemas digitais e dispositivos móveis;
- Suporte aos dialetos regionais;
- Sugestões para a representação das informações não literais;
- A especificação, projeto e implementação do sistema de geração e codificação da Língua de Sinais Brasileira;
- A realização de testes com o sistema desenvolvido;
- O levantamento de possíveis melhorias no sistema.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] RUSSELL, S.; NORVIG, P.. **Inteligência Artificial**. 2a ed. ed. Rio de Janeiro: Editora Campus, 2004.
- [2] B. Rime and L. Schiaratura. *Gesture and speech*, pages 239-281. Editions de la Maison des Sciences de l'Homme, 1991.
- [3] STUMPF, M. R.. **Língua de Sinais**: escrita dos surdos na Internet. V Congresso Ibero-Americano de Informática na Educação – RIBIE. Viñadelmar, Chile: [s.n.]. 2000. p. 1-8.
- [4] HANIM, 2011. Disponível em <<http://www.seamless3d.com/hanim/index.html>>. Acesso em: 11 Outubro 2011.
- [5] WHO - Deafness and hearing impairment. **World Health Organization**, 2010. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs300/en/>>. Acesso em: Outubro 2011.
- [6] 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.
- [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] GOES, M. C. R.. *Linguagem, Surdez e Educação*. Campinas: Editora Autores Associados, 1996.
- [9] 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.
- [10] BATISTA, C. E. C. F., ARAÚJO, T. M. U. de, OMAIA, D., ANJOS, T. C. dos, CASTRO, G. M. L. de, BRASILEIRO, F. V., SOUZA FILHO, G. L. de. “TVGrid: A Grid Architecture to use the idle resources on a Digital TV network”. Seventh IEEE International Symposium on Cluster Computing and the Grid — CCGrid, 2007.
- [11] ISO (2000). ISO/IEC 13818-1. Information technology – Generic coding of moving pictures and associated audio information - Part 1: Systems.
- [12] ISO (1998). ISO/IEC 13818-6. Information technology – Generic coding of moving pictures and associated audio information - Part 6: Extensions for DSM-CC.

- [13]ISO/IEC (1996), ‘So/iec13818-1 tr information technology - generic coding of moving pictures and associated information: part 1: systems’.
- [14]LEITE, L. E. C., SOUZA FILHO, G. L. de. “Recomendações para o Modelo de Referência: MIDDLEWARE (RFP04/2004) FlexTV”. Disponível em: <<http://www.tvdigital.inf.br/pdfs/relatorios/Middleware-UFPB.pdf>>. Acesso em: 12 Outubro de 2011.
- [15]GEM (2002) ETSI – European Telecommunications Standards Institute. Globally Executable MHP (GEM). TS 102 819, Novembro.
- [16]SOARES, L. F. G. “MAESTRO: The Declarative Middleware Proposal for the SBTVD”. In: 4th European Interactive TV Conference, 2006, Atenas. Proceedings of the 4th European Interactive TV Conference, 2006.
- [17]MARTINS, R. B., “Sistema Brasileiro de TV Digital: Metodologia para Escolha do Modelo de Referência”. Disponível em: <<http://www.telebrasil.org.br/arquivos/tvdigital.pdf>>. Acesso em: 14 Novembro de 2011.
- [18]BRITO, L. F.. Por uma gramática de língua de sinais. Rio de Janeiro: Editora Tempo Brasileiro, 1995.
- [19]QUADROS, R.M.; KARNOPP, L.B. *Língua de Sinais Brasileira: estudos lingüísticos*. Porto Alegre: Artmed, 2004.
- [20]DENARDI, R. M.; MENEZEZ 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.
- [21]Curso Básico da LIBRAS (Língua Brasileira de Sinais). Elaborado por Jonas Pacheco, Eduardo Estruc e Ricardo Estruc - V.8.08, 2008. Disponível em < www.surdo.org.br >.
- [22]FUSCO, E.. X-LIBRAS: Um ambiente virtual para a língua brasileira de sinais. Centro Universitário "Eurípedes de Marília" - UNIVEM. Marília, p. 158. 2004.
- [23]Lee, D.G., Fels, D.I., and Udo, J.P. (2007). Emotive Captioning. ACM Comput, Entertaint. v. 5, n. 2, p. 3-15 Article 11 (August 2007).
- [24]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.
- [25]HONG, R.; WANG, M.; XU, M.; YAN, S.; CHUA, T. Dynamic Captioning: Video Accessibility Enhancement for Hearing Impairment. Department of ECE, National University of Singapore. Singapore, p. 1-10, 2010.

- [26]“MobileASL,”. Disponível em: <<http://mobileasl.cs.washington.edu/>>. Acesso em: 26 Novembro de 2011.
- [27]“Visual Communications Lab”. Disponível em: < <http://foulard.ece.cornell.edu/>>. Acesso em: 26 Novembro de 2011.
- [28]CHON, J.; CHERNIAVSKY, N.; RISKIN, E.; LADNER, R. Enabling Access Through Real-Time Sign Language Communication Over Cell Phones. University of Washington, Seattle, p. 1-5, 2009.
- [29]SCHUMEYER, R.; HEREDIA, E.; BARNER, K. Region of Interest Priority Coding for Sign Language Videoconferencing. IEEE, Applied Science and Engineering Laboratories. Indianapolis, p. 1-6, 2005.
- [30]AGRAFIOTIS, D.; CANAGARAJAH, N.; BULL, D.; KYLE, J.; SEERS, H.; DYE, M. A Video Coding System for Sign Language Communication at Low Bit Rates. International Conference on Image Processing (ICIP). Dept. of Electrical & Electronic Engineering, University of Bristol, UK, p. 1-4, 2004.
- [31]Rybená, 2010. Disponível em: <<http://www.rybena.com.br/default/index.jsp>>. Acesso em: 19 Junho 2012.
- [32]AMORIM, M. L. C.; AL., E. RybenáTV: Solução para acessibilidade de surdos para TV Digital. Simpósio Brasileiro de Sistemas Multimídia e Web - Webmedia 2010. [S.l.]: [s.n.]. 2010. p. 243-248.
- [33]Mo_at A (1990) Implementing the PPM data compression scheme. IEEE Trans Commun, 38:1917-1921.
- [34]Cleary JG, Witten IH (1984) Data compression using adaptive coding and partial string matching. IEEE Trans Commun 32:396-402.
- [35]Batista LV, Meira MM (2004) Texture classification using the Lempel-Ziv-Welch Algorithm. Lect Notes in Comput Sci, 3171:444-453.
- [36]Bratko A et al (2006) Spam filtering using statistical data compression models, J Mach Learn Res 7:2673-2698.
- [37]Mahoui M, Teahan WJ, Sekhar WJT, Chilukuri S (2008) Identification of gene function using prediction by partial matching (PPM) language models, Proc 17th ACM Conf Inf Knowl Manag, 1:779-786.
- [38]Medeiros TFL, Cavalcanti, AB, et al (2011), Heart arrhythmia classification using the PPM algorithm, Proc Biosignals Biorobotics Conf, 1:1-5.

- [39] Santos GS, Silveira MS, Aluisio SM (2009) Produção de textos paralelos em língua portuguesa e uma interl__ngua de LIBRAS (Production of parallel texts in Portuguese and a LIBRAS interlanguage), Proc XXXVI Semin Integr Softw Hardw (SEMISH), 1:371-385.
- [40] BLAKOWSKI, G. and STEINMETZ, R. (1996): A media synchronization survey: reference model, specification and case studies. IEEE Journal on Selected Areas in Communications 14(1):5-35.
- [41] 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.
- [42] GingaCDN – Code Development Network. Disponível em: < <http://gingacdn.lavid.ufpb.br/> >. Acesso em: 14 Novembro 2011.
- [43] Blender, 2011. Disponível em: <<http://www.blender.org/>>. Acesso em: 14 Novembro 2011.
- [44] Educação Especial: Alfabetização de Surdos< <http://www.artigonal.com/educacao-infantil-artigos/educacao-especial-alfabetizacao-de-surdos-4939205.html> > Acesso em: 28 Julho 2012.
- [45] Jain, R. (1991), The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling, Willey, Nova York, EUA.
- [46] Gallo B et al (2009) Speech into sign language statistical translation system for deaf people. IEEE Latin America Trans, 7:400-404.
- [47] San-Segundo R et al (2008) Speech to sign language translation system for spanish. Speech Commun 50:1009-1020.
- [48] San-Segundo R et al. (2011) Design, development and _eld evaluation of a spanish into sign language translation system. Pattern Anal Applic. doi:10.1007/s10044-011-0243-9.
- [49] 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.
- [50] FGV/ABERT (2012), ‘Pesquisa sobre tv digital no brasil’. URL: <http://www.abert.org.br/site/images/stories/pdf/TVProgramacao.pdf>
- [51] M. Kipp M, et al., “Assessing the deaf user perspective on sign language avatars”, Proc 13th Intern ACM Conf Assist Technol, 2011, pp. 1-8.
- [52] R. San-segundo, et al., “Development, design and field evaluation of a Spanish into sign language translation system”, Pattern Anal Applic. doi:10.1007/s10044-011-0243-9.

ANEXO A – QUESTIONÁRIOS DE AVALIAÇÃO

LibrasTV: 1ª ETAPA DE TESTES

Como você classificaria esse trecho de vídeo considerando a sua compreensão e o uso correto de LIBRAS?

1. Os sinais são incorretos ou muito confusos para que se entenda o significado;
2. É difícil compreender, mas consegui entender o espírito da coisa;
3. Compreendo, mas está um pouco incorreto;
4. Compreendo e está correto;

Como você classificaria esse trecho de vídeo, como uma tradução do português:

1. A tradução está completamente incorreta;
2. Os conceitos básicos estão corretos, mas a maioria dos conceitos está incorreto ou tem informação faltando;
3. Boa tradução, poucas coisas incorretamente traduzidas ou faltando;
4. Excelente tradução, sem erros;

Como você classificaria esse trecho de vídeo com base na naturalidade das animações:

1. Os sinais não possuem nenhuma naturalidade;
2. Os sinais são muito pouco naturais;
3. Os sinais são razoavelmente naturais, similares a um intérprete humano;
4. Não existem diferenças em relação a um intérprete de LIBRAS;

Você acha que seria possível usar o sistema na ausência de um intérprete humano?

1. Não, de forma alguma;
2. Dificilmente, existem muitos prejuízos com relação ao entendimento;
3. Daria para usar, mas com algumas restrições;
4. Sim, daria para usar facilmente o sistema

Qual é, em sua opinião, o grau de relevância do sistema proposto?

1. Nada relevante, não modifica em nada a situação atual;
2. Pouco relevante, representa uma pequena melhoria para o estado atual;
3. Bastante relevante;
4. Atende completamente minhas necessidades.

Qual seria, em sua opinião, o principal problema do sistema proposto?

1. A qualidade dos sinais é ruim;
2. A qualidade da tradução não é boa;
3. Não aprovo o uso de animações (avatar);
4. Não existem problemas relevantes;

Como você caracteriza a velocidade de apresentação dessa animação?

1. O avatar apresenta os sinais de forma muito lenta;

2. O avatar apresenta os sinais de forma muito rápida;
3. A velocidade da animação está razoável, poderia ser mais lenta;
4. A velocidade da animação está razoável, poderia ser mais rápido;
5. A velocidade da animação está boa;

Como você classificaria o seu entendimento na identificação do personagem através da inclinação do avatar?

1. Não consegui identificar quem estava falando
2. Identifiquei algumas vezes quem estava falando
3. Demorei, mas consegui identificar quem estava falando
4. Entendi perfeitamente quem estava falando

Como você classificaria o seu entendimento da identificação do personagem através do uso de imagens com números e cores diferentes?

1. Não consegui identificar quem estava falando
2. Identifiquei algumas vezes quem estava falando
3. Demorei, mas consegui identificar quem estava falando
4. Entendi perfeitamente quem estava falando

Como você classificaria o seu entendimento a cerca da emoção sentida pelos personagens?

1. Não consegui entender nada do que o personagem estava sentindo
2. Compreendi muito pouco as emoções sentidas pelos personagens
3. Compreendi as emoções, mas não as respectivas intensidades
4. Entendi perfeitamente as emoções e suas respectivas intensidades

Como você classificaria o nível de entendimento a cerca da imagem onde demonstrava palmas

1. Não compreendi nada
2. Compreendi muito pouco o significado da imagem
3. Entendi que eu teria que bater palma
4. Compreendi perfeitamente que existiam sons de palmas

Se existisse um canal onde ensinasse o significado de cada pista visual que aparece perto da animação, qual seria seu nível de entendimento?

1. Ruim
2. Pouco
3. Bom
4. Perfeito

LibrasTV: 2ª ETAPA DE TESTES

INFORMAÇÃO USUÁRIO:

SEXO: () MASCULINO () FEMININO

IDADE: _____

VOCÊ É? () SURDO () INTÉRPRETE () INSTRUTOR

COMO VOCÊ SURDO?

ESCOLHER:

- () SURDO NASCENÇA
- () TER ALGUÉM FAMÍLIA TAMBÉM SURDO
- () SUA FAMÍLIA USAR LIBRAS SUA CASA
- () VOCE PARTICIPAR COMUNIDADE SURDO?
- () TRABALHA ALGUMA ESCOLA PARA SURDO?
- () ESTUDAR ESCOLA PARA SURDOS?
- () ESTUDAR ESCOLA JUNTO OUVINTE?

VOCÊ SABER LER? () SIM () NÃO

QUAL SEU GRAU ESCOLARIDADE?

- () NUNCA ESTUDAR
- () ENSINO FUNDAMENTAL INCOMPLETO (AINDA NÃO TERMINAR 5º AO 9º ANOS).
- () ENSINO FUNDAMENTAL COMPLETO (TERMINAR 5º AO 9º ANOS).
- () ENSINO MÉDIO INCOMPLETO (AINDA NÃO TERMINAR 1º AO 3º ANOS).
- () ENSINO MÉDIO COMPLETO (TERMINAR 1º AO 3º ANOS).
- () ENSINO SUPERIOR INCOMPLETO (NÃO TERMINAR UNIVERSIDADE).
- () ENSINO SUPERIOR COMPLETO (CONCLUIR CURSO UNIVERSIDADE).
- () OUTRO: _____

ESCOLHER NÍVEL VOCE TER CONHECIMENTO PORTUGUÊS (6 - avançado, 1 - não sei nada) ?

- () 1 () 2 () 3 () 4 () 5 () 6

LIBRAS ESCOLHER NÍVEL VOCE TER CONHECIMENTO LIBRAS (6 - avançado, 1 - não sei nada) ?

- () 1 () 2 () 3 () 4 () 5 () 6

AVALIAR VÍDEO 1 (LUZIA E A VACA ANDORINHA):

1) FÁCIL ENTENDER? (6 - CLARO); (1 - CONFUSO)

- () 6 () 5 () 4 () 3 () 2 () 1

2) SEGUE GRAMÁTICA LIBRAS? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

3) SINALIZAÇÃO PARECER NATURAL/NORMAL? (6 - PARECER PESSOA); (PARECER ROBÔ)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

4) VELOCIDADE APRESENTAÇÃO AVATAR? (6 - MUITO RÁPIDO); (1 - MUITO LENTA)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

5) MOVIMENTOS MÃOS ESTÁ CORRETO? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

6) EXPRESSÃO FACIAL ESTÁ CORRETO? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

7) VÍDEO MOSTRAR DISCUSSÃO CASAL. O-QUE DISCUSSÃO?

- ☐ CASAL DISCUTIR PROBLEMA SOBRE FIDELIDADE OU CIÚME.
- ☐ CASAL DISCUTIR PROBLEMA SOBRE DINHEIRO.
- ☐ CASAL DISCUTIR PROBLEMA SOBRE FILHOS
- ☐ SEI-NÃO.

8) POR QUE CASAL DISCUTIR?

- ☐ PORQUE SÓ TER CAFÉ.
- ☐ PORQUE ONTEM ESPOSA CHEGAR MADRUGADA.
- ☐ PORQUE FILHO DELES NÃO ACORDAR E NÃO IR ESCOLA.
- ☐ SEI-NÃO.

9)QUAL LUGAR MARIDO AVISAR ESPOSA IR PROBLEMA RESOLVER?

- ☐ IR MERCADO OU MERCEARIA
- ☐ IR PSICOLOGO OU MÉDICO
- ☐ IR ADVOGADO OU JUSTIÇA
- ☐ SEI-NÃO

10)MARIDO MANDAR ESPOSA IR MERCEARIA,MAS ESPOSA FALAR PARA MARIDO NÃO DAR CERTO PORQUE?

- ☐ PORQUE MÉDICO PODER-NÃO ATENDER.
- ☐ HOMEM MERCADO FALAR NAO VENDER FIADO.
- ☐ ADVOGADO SER CARO.
- ☐ SEI-NÃO

11)MARIDO SAIR XATEADO O-QUE ELE FALAR ESPOSA FINAL HISTÓRIA VÍDEO?

- ☐ “PODER-NÃO SAIR TRABALHAR ATRASADO.”
- ☐ “PRECISAR IR MÉDICO TODO JEITO.”
- ☐ “PODER-NÃO FICAR FOME E IR TRABALHAR.”

☐ SEI-NÃO.

AVALIAR VÍDEO 2 (GLOBO - JORNAL NACIONAL):

1) FÁCIL ENTENDER? (6 - CLARO); (1 - CONFUSO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

2) ACOMPANHA GRAMÁTICA LIBRAS? (6 - PERFEITO); (1- MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

3) SINALIZAÇÃO PARECER NATURAL/NORMAL? (6 - PARECER PESSOA); (PARECER ROBÔ)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

4) VELOCIDADE APRESENTAÇÃO AVATAR? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

5) MOVIMENTOS MÃOS ESTÁ CORRETO? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

6) EXPRESSÃO FACIAL ESTÁ CORRETO? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

7)O-QUE CATÁSTROFE APARECER REPORTAGEM ACONTECER?

☐ TERREMOTO E TREMOR TERRA

☐ TEMPESTADE OU MUITA CHUVA

☐ TSUNAMI OU ONDA GIGANTE

☐ SEI-NÃO

8) QUANTOS CORPOS ENCONTRAR SOTERRADO (SIGNIFICA: CORPO EMBAIXO TERRA)?

☐ 1

☐ 2

☐ 5

☐ SEI-NÃO

9) O-QUE ACONTECER CASA PESSOAS MORRER?

☐ CASA PEGAR FOGO.

☐ AGORA OUTRAS PESSOAS MORAR LÁ.

☐ CHUVA DESABAR CASA.

☐ SEI-NÃO.

AVALIAR VÍDEO 3 (TV RECORD - HOJE EM DIA):

1) FÁCIL ENTENDER? (6 - CLARO); (1 - CONFUSO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

2) ACOMPANHA GRAMÁTICA LIBRAS? (6 - PERFEITO); (1- MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

3) SINALIZAÇÃO PARECER NATURAL/NORMAL? (6 - PARECER PESSOA); (PARECER ROBÔ)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

4) VELOCIDADE APRESENTAÇÃO AVATAR? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

5) MOVIMENTOS MÃOS ESTÁ CORRETO? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

6) EXPRESSÃO FACIAL ESTÁ CORRETO? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

7) QUAL TEMA DUAS MULHER DISCUTIR COMEÇO VÍDEO?

☐ ECONOMIA

☐ NUTRIÇÃO

☐ MODA

☐ SEI-NÃO

8) O-QUE APRESENTADORA GIANNE ALBERTONI FALAR VÍDEO?

☐ NÃO COMER MUITA MASSA.

☐ PODER USAR POUCO BRILHO.

☐ USAR MUITO BRILHO ROUPA.

☐ SEI-NÃO.

9) QUAL RECEITA EDU GUEDES FAZER VÍDEO?

☐ DOCE PÊSSEGO

☐ DOCE MAMÃO

☐ MOUSSE MARACUJÁ

☐ SEI-NÃO

10) MULHER NOME “MELISSA” JUNTO EDU GUEDES FAZER RECEITA, MAS ELA TAMBÉM TRABALHAR E O-QUE FAZER MAIS?

☐ MELISSA PROFESSORA

☐ MELISSA TRABALHAR E CUIDAR ANIMAIS

☐ MELISSA VENDER ROUPAS.

☐ SEI-NÃO.

AVALIAR VÍDEO 4 (SINTELL):

1) FÁCIL ENTENDER? (6 - CLARO); (1 - CONFUSO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

2) ACOMPANHA GRAMÁTICA LIBRAS? (6 - PERFEITO); (1- MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

3) SINALIZAÇÃO PARECER NATURAL/NORMAL? (6 - PARECER PESSOA); (PARECER ROBÔ)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

4) VELOCIDADE APRESENTAÇÃO AVATAR? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

5) MOVIMENTOS MÃOS ESTÁ CORRETO? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

6) EXPRESSÃO FACIAL ESTÁ CORRETO? (6 - PERFEITO); (1 - MUITO POUCO)

☐ 6 ☐ 5 ☐ 4 ☐ 3 ☐ 2 ☐ 1

7)MULHER TER ANIMAL ESTIMAÇÃO, QUAL?

☐ GALINHA

☐ GATO

☐ DRAGÃO

☐ SEI-NÃO

8)O-QUE GUARDIÃO FALAR MULHER?

☐ FALAR ELA PERTO ENCONTRAR PRÓPRIO ANIMAL.

☐ FALAR ELA PROCURAR IR CAMINHO ERRADO E ESTAR LONGE.

☐ FALAR ELA SER FILHA DELE.

☐ SEI-NÃO

9) QUAL SER HISTÓRIA LANÇA PERSONAGEM USAR?

☐ LANÇA TER PASSADO TRISTE. MATAR MUITO INOCENTE.

☐ LANÇA TER GLÓRIA PASSADO. MATAR MUITO BANDIDO.

☐ LANÇA PASSADO PERTENCER GRANDE GUERREIRO HÉRCULES PRÓPRIO.

☐ SEI-NÃO

10)O-QUE SIGNIFICAR SINTELL?

☐ SIGNIFICAR TERRA PERIGOSA.

☐ SIGNIFICAR TERRA FRIO

☐ SIGNIFICAR TERRA DRAGÃO

☐ SEI-NÃO

IDENTIFICAÇÃO DO PERSONAGEM

1) QUAL DOIS FALAR: NÃO-PODER SAIR COM-FOME DEPOIS IR TRABALHAR?

- ☐) HOMEM
- ☐) MULHER
- ☐) HOMEM MULHER
- ☐) SEI-NÃO

2) QUAL DOIS FALAR: NADA COMER NÃO-TER FARINHA?

- ☐) HOMEM
- ☐) MULHER
- ☐) HOMEM MULHER
- ☐) SEI-NÃO.

3) QUAL DOIS FALAR: IR DEPOIS MERCEARIA?

- ☐) HOMEM
- ☐) MULHER
- ☐) HOMEM MULHER
- ☐) SEI-NÃO.

4) QUAL DOIS PERGUNTAR: SÓ CAFÉ?

- ☐) HOMEM
- ☐) MULHER
- ☐) HOMEM MULHER
- ☐) SEI-NÃO.

5) SE NÃO-TER INTÉRPRETE-PESSOA, VOCÊ ACEITAR SUBSTITUIR "INTÉRPRETE-VIRTUAL" ? (6 - PERFEITO); (1 - MUITO POUCO)

☐) 6 ☐) 5 ☐) 4 ☐) 3 ☐) 2 ☐) 1

6) VOCÊ CONSEGUIR ENTENDER QUAL PESSOA FALAR? (6 - PERFEITO); (1 - MUITO POUCO)

☐) 6 ☐) 5 ☐) 4 ☐) 3 ☐) 2 ☐) 1

ANEXO B – PRODUÇÃO TÉCNICA E CIENTÍFICA

Como resultado do desenvolvimento deste trabalho, também foi gerada uma produção técnica e científica, incluindo patentes registradas, publicações em periódicos e conferências nacionais e internacionais. O conjunto de produções é detalhado abaixo:

- **Dois Artigos Aceitos para Publicação em Periódico:**

1. Ferreira, F. L. S., **F. H. Lemos**, G. L. Souza Filho, et al. (2012), 'Generating Window of Sign Languages on ITU J.200-Based Middlewares', International Journal of Multimedia Data Engineering and Management 3(2), 20-40.
2. Araújo, T. M. U., **F. H. Lemos**, G. L. Souza Filho, et al. (2012), 'Automatic Generation of Brazilian Sign Language Windows for Digital TV Systems', Journal of the Brazilian Computer Society (Aceito para Publicação).

- **Dois Artigos Completos Publicados em Conferências Nacionais e Internacionais:**

1. Ferreira, F. L. S., **F. H. Lemos**, G. L. Souza Filho, et al. (2011), Providing Support for Sign Languages in Middlewares Compliant with ITU J.202, em 'Proceddings of the 2011 IEEE International Symposium of Multimedia, ISM'11', Dana Point, EUA, pp. 149-156.
2. **Lemos, F. H.**, T. M. U. Araújo, G. L. Souza Filho, et al. (2011), Uma Proposta de Protocolo de Codificação de LIBRAS para Sistemas de TV Digital, em 'Anais do XVII Simpósio Brasileiro de Sistemas Multimídia e Web, Webmedia' 11', Florianópolis, Brasil, pp. 103-110.

- **Um Prêmio de Melhor Artigo em Conferência Internacional:**
 1. Best paper award runner-up no IEEE International Symposium on Multimedia (ISM2011), IEEE International Symposium on Multimedia (ISM2011).
- **Uma Patente Depositada**
 2. **Lemos, F. H.,** G. L. Souza Filho, et al. (2009), Sistema para Geração Automática de Legendas em LIBRAS em Cinemas Digitais, 2012,