

Estruturação de registro de Libras em computador, para reconhecimento de sinais e animação de Avatares

Victoria M T Palhares

Marcelo Krieger

Stefano La Porta

Internal Research Reports

Number 49 | September 2016

Estruturação de registro de Libras em computador, para reconhecimento de sinais e animação de Avatares

Victoria M T Palhares

Marcelo Krieger

Stefano La Porta

CREDITS

Publisher:

MAXWELL / LAMBDA/CCPA/VRac

Sistema Maxwell / Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos

<http://www.maxwell.vrac.puc-rio.br/>

Organizers:

Alexandre Street de Aguiar

Delberis Araújo Lima

Cover:

Ana Cristina Costa Ribeiro

Estruturação de registro de Libras em computador, para reconhecimento de sinais e animação de Avatares.

Victoria M. T. Palhares¹, Marcelo Krieger², Stefano La Porta³.

¹ Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, 55-21-3527-1001, victoriapalhares@aluno.puc-rio.br.

² Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, 55-21-3527-1001, mkrieger@esp.puc-rio.br.

³ Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, 55-21-3527-1001, slaporta@esp.puc-rio.br.

Resumo- O objetivo do Projeto TraLP é o desenvolvimento de um aplicativo capaz de traduzir Libras para Português e Português para Libras, para as plataformas Android, Windows Phone e iOS. O intuito da ferramenta é facilitar a conversação entre falantes de português e alfabetizados em Libras, usando os recursos existentes nos dispositivos: câmera para captura de sinais e tela para apresentação dos sinais através de um Avatar. Para tanto, é necessária uma representação sintética, que permita o armazenamento e identificação destes sinais para tradução, levando em consideração as limitações de memória dos aparelhos. Baseado na análise da literatura atual e nos resultados das pesquisas mais recentes sobre parametrização de Libras, foi possível a decomposição dos sinais da língua em seus componentes fundamentais: configuração de mão, ponto de articulação, orientações de palma, movimentos e expressões faciais. Este trabalho apresenta o modelo formulado da solução para descrição deste vocabulário e seu armazenamento eficiente no banco de dados.

Palavras-chave— Libras; Smartphone; Armazenamento; Otimização.

1 Introdução

No momento em que se decide criar uma ferramenta que traduza uma língua para outra, com origem e configurações completamente diferentes, a exemplo de Português e Mandarim, vê-se que não é uma tarefa tão simples. A transformação de um idioma para o outro não pode ser feita de maneira direta, muitas vezes precisando de modificações em sua estrutura e de adaptações. A partir do momento que estamos lidando com uma língua falada e com outra gestual, como a linguagem dos deficientes auditivos, o caso é bem mais complexo, exigindo mudanças na estrutura morfológica e sintática, [6].

O Projeto TraLP tem o objetivo de suprir a dificuldade de comunicação entre deficientes auditivos e ouvintes, uma vez que a maioria das pessoas não conhece Libras. A ferramenta traduzirá Libras para Português e Português para Libras, nas plataformas Android, Windows Phone e iOS. No âmbito deste projeto, é necessário construir um dicionário Português-Libras. Para a versão Libras do dicionário, existe necessidade de armazenar os muitos sinais que compõem o vocabulário de Libras. Em um primeiro momento, é natural cogitar a ideia de

gravar um vídeo para cada uma das palavras. É trabalhoso, porém simples. Todavia, este caminho geraria mais de três mil vídeos para compor um vocabulário básico. Se considerarmos vídeos de 10 segundos cada, esta abordagem demandaria mais de 8 horas de gravações, as quais, mesmo altamente compactadas, demandariam cerca de 1,5 GB de memória. Esta demanda é superior ao espaço de armazenamento da maior parte dos dispositivos móveis disponíveis nos dias atuais. O objetivo deste trabalho é apresentar uma formulação alternativa, mais econômica do ponto de vista de consumo de memória, viabilizando o desenvolvimento do projeto.

2 Parametrização de Libras

Da mesma forma que os idiomas escritos são representados pela utilização de componentes básicos chamados grafemas e as línguas faladas são desenvolvidas a partir de fonemas, Libras também pode ser decomposta em um conjunto de parâmetros básicos, que combinados, geram os sinais representativos das ideias que devem ser comunicadas, [3], [4]. De acordo com as mais recentes pesquisas [1], [2] sobre parametrização de Libras, os sinais da língua podem ser decompostos em quesitos como configurações de mão, pontos de articulação, orientações de palma, movimentos e expressões faciais.

Dentro disto, foram identificadas 74 configurações de mão que, combinadas aos outros quesitos, formam os sinais. Algumas delas, por si sós, já representam letras.

Para os pontos de articulação, que indicam uma referência de onde é feito o sinal, foram identificados 82 pontos no total, a exemplo de “na frente do tórax” e “acima da cabeça”. O mesmo sinal, em pontos de articulações diferentes, pode significar coisas completamente distintas. Um exemplo disso é o sinal com a mão completamente fechada. Posicionada tocando a testa significa o verbo “aprender”, Fig. 1. Já tocando o queixo significa “sábado”, Fig. 2.

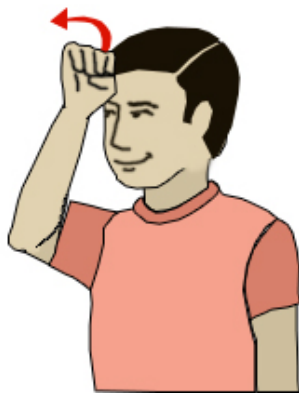


Figura 1. Sinal em Libras para “aprender”



Figura 2. Sinal em Libras para “sábado”

Em relação à orientação de palma, ou direção, são adotados tipos como “para baixo”, “para cima”, “para esquerda”, “para direita”, “para dentro” e “para fora”, muito importantes no momento em que são combinadas com a configuração de mão e com o ponto de articulação.

Quanto aos movimentos, foram identificados os seguintes tipos: retilíneo (em formato de reta), helicoidal (em formato de hélice), circular (em forma de círculo), semicircular (em forma de meia lua), sinuoso (na forma de um conjunto de curvas) e angular (em zigue-zague). Também estão inclusos movimentos de dedos como “uni-los pelas pontas”, de corpo como “inclinar para frente e para trás” e de mãos como “balançar para cima e para baixo”.

Adicionalmente, ainda relacionadas aos movimentos dos sinais, existem duas características adicionais: a frequência e a intensidade do sinal. Quando um sinal apresenta movimentos repetitivos, a frequência é aplicada. Já a segunda característica é usada em substituição aos advérbios de intensidade. Movimentos suaves ou fortes suprem essa lacuna.

Por fim, relacionada às expressões faciais, os mais diversos sentimentos são expressos pelos movimentos do rosto como tristeza, irritação, surpresa, alegria e remorso. No momento que vai se comunicar em Libras, o deficiente auditivo utiliza seu rosto como instrumento para demonstrar o seu sentimento em relação ao acontecimento ou ação. Em uma frase como “Assisti o filme Titanic”, ele pode tanto expressar tristeza, quanto felicidade, personalizando o discurso.

Estes elementos em conjunto compõem um sinal e em seguida uma mensagem em Libras. Como foi dito anteriormente, pontos de articulação distintos com mesma

configuração de mão, representam sinais distintos. Na mesma linha de raciocínio, movimentos em ordens diferentes representam mensagens diferentes, assim como a alteração da ordem das palavras em uma frase em português atribui-lhe um significado distinto.

3 Modelagem para Armazenamento

Pelo que foi descrito anteriormente, deve-se estruturar um banco de dados sintético que permita o armazenamento do dicionário Português-Libras. Apenas para reforçar, o dicionário isoladamente não é suficiente para que toda a tradução seja feita, mas sua existência é imprescindível para que a finalidade do projeto seja atingida.

A estrutura básica do dicionário é a palavra (conceito que inclui letras e números). Em Libras, uma palavra é composta por um ou mais sinais. Cada sinal parte de uma configuração de mão, ponto de articulação, orientação de palma e expressão facial. Em alguns casos, isto é suficiente para descrever um sinal. Em outros, um ou mais movimentos alteram este ponto de partida e complementam o sinal.

Para a configuração de mão, um único byte é capaz de armazenar cada uma das possibilidades, uma vez que um byte é capaz de armazenar 256 números diferentes. O mesmo acontece com o ponto de articulação.

Outro aspecto tratado foi a orientação da palma da mão. Como foi dito no capítulo anterior, existem seis orientações comumente encontradas. Pesquisas adicionais, [5], demonstram que existem mais alternativas possíveis. Para evitar a complicação de antecipar cada uma delas, optou-se por descrever a orientação de palma a partir de três ângulos, obtendo uma formulação mais simples, eficiente e geral. Considerando uma posição inicial da mão, sempre voltada pra baixo, com os dedos apontando para frente e trabalhando com as três angulações, é possível achar a orientação de palma desejada para o início do sinal, Fig. 3. Temos então um ângulo I, entre o eixo x e y, ângulo II entre o eixo x e z e um ângulo III em que a mão gira sobre o seu próprio eixo (y e z). Com esta solução podemos não só descrever as mais básicas direcionalidades citadas anteriormente, como também as mais complexas, por exemplo, em que a mão faz ângulo de 45 graus no ângulo II e é rotacionada de 90 graus no ângulo III.

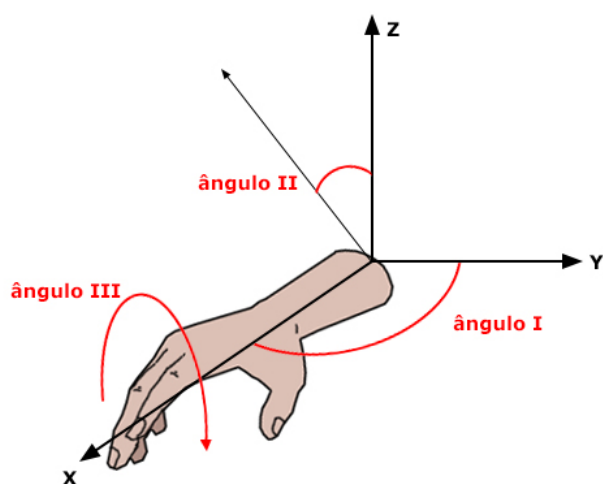


Figura 3. Esquemática da orientação de palma

O passo seguinte consistiu em registrar e descrever os movimentos. Percebemos então que seria possível descrevê-los a partir de quatro tipos básicos: linear, circular, de mudança de configuração de mão e rotacional.

O primeiro indica uma trajetória feita pela mão, em uma linha reta, em qualquer direção no espaço. Há, então, como parâmetros da descrição do movimento, um ângulo 1 e 2 para identificar sua direção, e o tamanho da reta a ser percorrida. Assim, podemos descrever os movimentos, retilíneo e angular, já vistos no estudo.

O movimento circular corresponde ao deslizamento da mão sobre um círculo máximo de uma superfície esférica. A composição de vários movimentos circulares sobre esferas de diferentes diâmetros permite sintetizar os movimentos circular, semicircular, helicoidal e sinuoso, citados anteriormente. Para descrevê-lo, temos como parâmetros o raio da esfera sobre a qual será feito o movimento e as coordenadas iniciais e finais do movimento, que podem ser expressas como uma latitude e longitude inicial e final, totalizando cinco parâmetros.

O movimento de mudança de configuração de mão serve para representarmos movimentos de dedos, como, por exemplo, uni-los pelas pontas. Para descrever este tipo, é necessário saber a configuração de mão inicial e final. Contudo, no momento de caracterizar um sinal, já registramos a configuração de mão inicial. Logo, quando o movimento é de mudança de uma configuração para outra, é necessário ter como único parâmetro a configuração de mão final.

Por último, o movimento rotacional indica uma rotação em torno do eixo paralelo ao dedo médio, eixo 1, ou atravessando o punho, eixo 2, como mostra a Fig. 4. Para este tipo, indica-se, então, o eixo em que está ocorrendo a rotação e o ângulo a ser feito, sendo estes os parâmetros necessários.

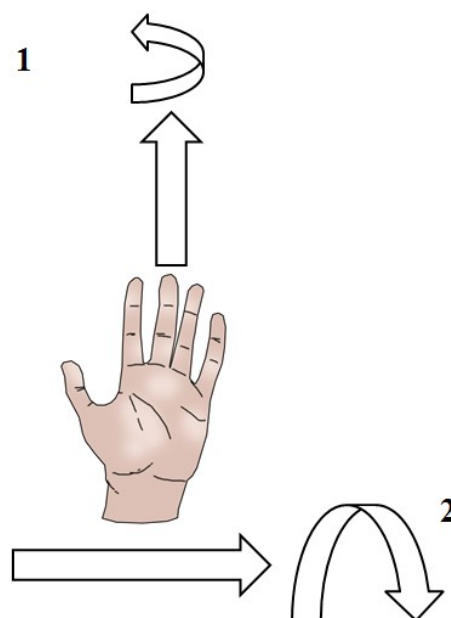


Figura 4. Sistema do movimento rotacional

A intensidade dos gestos é descrita por um parâmetro de tempo de execução do sinal, afinal, para quem vê, um gesto intenso costuma ser mais rápido do que o normal.

Em relação à frequência, basta que sejam registradas repetições na sequência de códigos dos movimentos.

No banco de dados criado, foi projetada uma tabela somente para a sistematização dos movimentos, Tabela 1, indicando os tipos, os parâmetros usados para cada tipo e os seus respectivos códigos.

Tabela 1. Movimentos

CODMOV	TIPO	PARAM1	PARAM2	PARAM3	PARAM4	PARAM5
1	LINEAR	ÂNGULO1	ÂNGULO2	TAMANHO		
2	CIRCULAR	LATITUDE1	LONGITUDE1	LATITUDE2	LONGITUDE2	RAIO
3	ROTACIONAL	ÂNGULO	EIXO			
4	MUDANÇA CM	CM FINAL				

A última característica estudada foi as expressões faciais. Ainda é muito difícil trabalhar e distinguir os mais diferentes sentimentos pelos movimentos do rosto. Ao fazer o reconhecimento de imagem, saber a localização exata dos olhos, da boca, do nariz e diferenciar se, por exemplo, a boca estando aberta, significa medo ou surpresa, é ainda muito complexo. Optou-se pela simplificação: uso de expressões básicas, como felicidade, tristeza, ou neutro, para que assim a interpretação tenha mais chances de sucesso. Adicionalmente, viu-se que não será possível trabalhar com um movimento de corpo, já que a câmera do dispositivo só enquadrará os membros superiores.

Segue o modelo entidade relacionamento (MER) final, Fig. 5, que engloba tanto as características como configuração de mão, orientação de palma, expressão facial e ponto de articulação, como também os movimentos. Cada uma das

entidades tem os seus respectivos parâmetros, essenciais para uma descrição precisa.

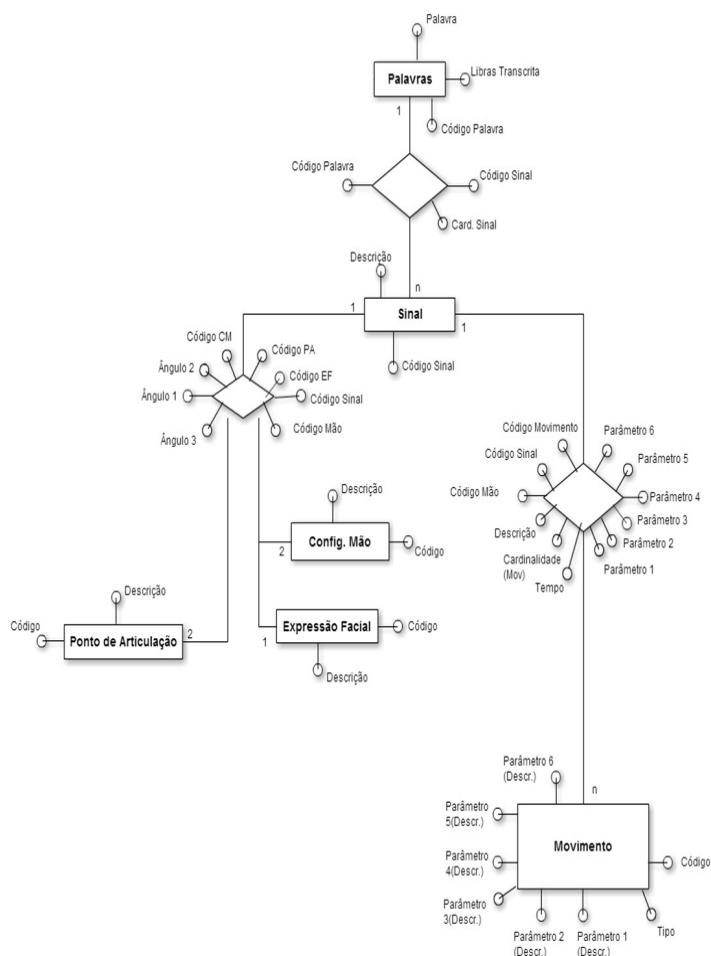


Figura 5. Modelo Entidade Relacionamento

Além das estruturas básicas resultantes das considerações anteriores, existe uma série de estruturas complementares, necessárias para implementar o modelo. Por exemplo, existe uma tabela que relaciona sinais e movimentos. Nesta, existe um campo “cardinalidade”, que indica a ordem da execução do movimento e se existem movimentos executados simultaneamente; tempo, para nos auxiliar na questão da intensidade do gesto a ser feito; código do sinal, para indicar qual deles está sendo relacionado; e código do movimento, para concluir esse relacionamento. . Da mesma forma, na execução de um sinal, é necessário indicar a que mão aplicam-se as configurações determinadas. Na tabela de Sinais X Movimentos abaixo, vê-se que há dois tipos de movimentos, rotacional e circular, para a letra J com a mesma cardinalidade. Isso indica que há uma rotação e uma trajetória com as mãos em um formato de semicírculo, simultaneamente.

A estrutura proposta permite o armazenamento de um sinal em menos de 1KB para o caso de um sinal sem movimento e até 10KB para sinais mais complexos. Desta forma, um dicionário de 3.000 vocábulos ocupa um espaço de cerca de 15MB, gerando uma economia da ordem de 100:1 em relação à memória necessária para armazenar filmagens ou animações dos sinais e viabilizando sua implementação nos dispositivos móveis.

4 Conclusão

No desenvolvimento de um aplicativo de tradução de Libras para Português e Português para Libras, surge a questão da construção de um dicionário Português-Libras e o consequente armazenamento do vocabulário da linguagem de sinais. Os dispositivos atuais apresentam limitações severas no espaço disponível, tornando inviável o armazenamento de vídeos ou animações individuais para cada um dos sinais que compõe o vocabulário do aplicativo. A partir dos componentes fundamentais de Libras, como configurações de mão, ponto de articulação, orientações de palma, movimentos e expressões faciais, foi modelada uma estrutura de armazenamento compacta e econômica. Nesta solução, é possível armazenar um vocabulário de tamanho significativo na memória dos dispositivos. O próximo passo é a geração dos algoritmos necessários para que, na animação e no reconhecimento dos gestos pela câmera, sejam compreendidos os códigos numéricos usados para cada característica do sinal. Desta forma, haverá uma aplicação eficiente, compacta e de nível satisfatório para os usuários.

Agradecimentos

Este trabalho foi desenvolvido com recursos do CNPq, oriundos da chamada MCTI-SECIS/CNPq No 84/2013 - Tecnologia Assistiva.

Referências

- [1]- Centro Universitário de Goiás, uni-Anhanguera. (2011). *Gramática da Língua de Sinais, Módulo IV*. Disponível em: <http://anhanguera.edu.br/ead/index.php/modulo/aula/103/1034/gramatica/5115>
- [2]- I. Fajardo, et al, “Mapeamento estruturado de Libras para utilização em sistemas de comunicação”, não publicado.
- [3]- Prime Cursos. *Curso de Libras Básico*; Disponível em: <http://www.primecursos.com.br/libras-basico/>
- [4]- Instituto Nacional da Educação de Surdos. (2008). *Dicionário da Língua Brasileira de Sinais*. Disponível em: <http://www.acessibilidadebrasil.org.br/libras/>
- [5]- A. M. S. Duarte, *Comunicando com as mãos*, Belo Horizonte, Brasil, Associação dos Surdos de Teófilo Otoni- MG, 1998.
- [6]- D. Sousa. (2010). *Um olhar sobre os aspectos linguísticos da Língua Brasileira de Sinais* (1ª ed)[Online]. Disponível em: <http://www.porsinal.pt/index.php?ps=artigos&idt=arte&cat=9&idart=129>