

---

# Sistema de Animação de Humanos Virtuais Voltado para o Ensino de Libras

Andréia R. de A. Schneider, Luciana P. Nedel

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{araschneider, nedel}@inf.ufrgs.br

**Abstract.** *Deaf people have a limited capacity of using oral language to communicate. Because of this, they use gestural languages as their native language. As these languages are dissociated from written languages and are massively based on gestures, the education on sign languages involves the use of a kind of resources normally not used in other languages education. This work proposes the development of a system to present Libras (Brazilian sign language) by animating virtual humans. The system is dedicated for the education of deaf and non-deaf people. The paper describes the system architecture, discuss and present the implementation of other important parameters involved in the movement generation: the enunciation space, the velocity, the time and the gesture amplitude. The results are presented and discussed.*

**Resumo.** *Os surdos possuem a capacidade de utilizar a língua oral para se comunicar limitada e por isso tem como língua materna as línguas gestuais. Devido à massiva utilização de gestos e da sua dissociação da escrita, o ensino de linguagens de sinais requer o suporte de recursos não usuais no ensino de outras línguas. Este trabalho propõe o desenvolvimento de um sistema de apresentação para Libras (a língua brasileira de sinais) através da animação de humanos virtuais, dedicado ao ensino deste tipo de linguagem para indivíduos surdos e não surdos. O artigo descreve não apenas a arquitetura do sistema, mas também discute e apresenta a implementação de outros parâmetros importantes na geração dos movimentos, como: espaço de enunciação, velocidade, tempo e amplitude dos gestos. Os resultados são apresentados e discutidos.*

## 1. Introdução

A forma mais comum de comunicação entre pessoas é a língua falada. Através dela se dá a transferência de conhecimento e o favorecimento da organização de idéias. Os surdos tem essa forma de comunicação tradicional afetada não tendo condições, portanto, de aprender a língua da comunidade ouvinte que o cerca. Isso faz com que os mesmos tenham como língua materna as línguas gestuais, que utilizam a visão como fonte receptora dos movimentos e os gestos corporais em sua formação. A falta dessa comunicação oral coloca o surdo à margem da sociedade ouvinte, tendo dificuldades de usufruir serviços básicos, como acesso a hospitais, correio e escolas, já que os ouvintes também tem dificuldades em aprender a língua gestual.

É fato que a animação de movimentos humanos através de humanóides está cada vez mais presente no cotidiano: em filmes, jogos e telejornais, sendo aliada importante

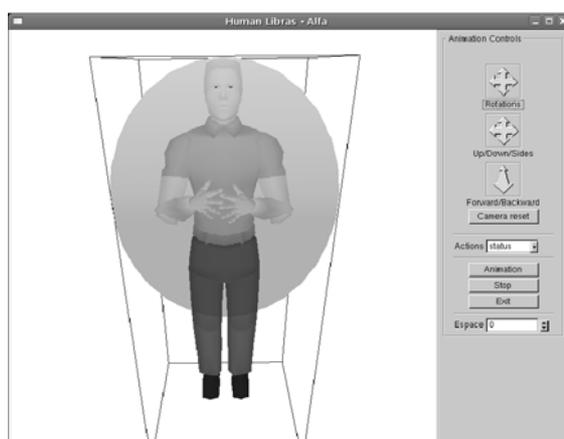
---

em diversas áreas, como na psicologia (tratamento de fobias)[Emmelkamp et al. 2002] e nos esportes (reprodução dos movimentos dos atletas)[Bideau et al. 2004]. É facilmente imaginável, portanto, a utilização dessas técnicas para auxiliar os surdos em tarefas do dia a dia, como pagamento de contas, visita a museus e até mesmo na alfabetização, fazendo assim com que os surdos tenham uma inclusão cada vez maior na sociedade ouvinte que o circunda. Ela também pode auxiliar ouvintes e não ouvintes a aprenderem língua de sinais.

As línguas de sinais são formadas pela associação de movimentos das partes acima do quadril. Elas são definidas como sendo gestuais-visuais porque são percebidas com o olhos e executadas com as mãos. Segundo Karnopp [Karnopp 2000], enquanto o emissor constrói uma sentença a partir de vários elementos (configurações de mãos, braços, ombro, expressão facial, etc.) o receptor utiliza os olhos ao invés dos ouvidos para entender o que está sendo comunicado. Desta forma, já que a informação linguística é recebida pelos olhos, os sinais são construídos de acordo com as possibilidades perceptuais do sistema visual humano.

Macarto [Marcato et al. 1974] afirma que para se ter uma representação computacional de Libras, deve-se combinar três parâmetros: a configuração das mãos (forma como estão posicionados os dedos e as mãos), o ponto de articulação (área no corpo ou no espaço de enunciação definido pelo corpo, em que ou perto do qual o sinal é articulado [Klima and Bellugi 1975]) e o movimento, além das componentes não manuais. Assim, para que uma animação seja gerada, alguns aspectos devem ser estudados, tais como o espaço de enunciação dos movimentos (espaço de configuração), a interpolação dos gestos, o aspecto temporal (tensão e velocidade do movimento) e as expressões faciais.

Espaço de enunciação é o local, no espaço tridimensional, aonde os sinais são articulados [de Quadros and Karnopp 2004]. A parte principal é a parte que é alcançada com os braços não muito estendidos (como quem tenta alcançar um objeto distante), nem muito dobrados (figura 1). Ele é um espaço ideal, levando-se em conta que os interlocutores estão de frente um para o outro. Porém podem haver situações em que esse espaço venha a ser totalmente redimensionado, como por exemplo, no caso em que os interlocutores estão fisicamente distantes, ou quando o espaço é muito restrito.



**Figure 1. Espaço de enunciação**

---

A passagem de um gesto para outro, assim como sua animação, deve ser feita de forma suave, sem que haja quebra de continuidade, fazendo assim com que o movimento seja mais realista. Como um gesto pode começar em uma locação e com uma configuração de mão, e terminar em outra e com uma configuração totalmente diferente, deve-se analisar qual a melhor forma de se fazer essa transição. Entende-se por locação a área do corpo, ou no espaço de enunciação, em que ou perto do qual o sinal é articulado [Klima and Bellugi 1975]).

Um movimento em Libras tem várias características que devem ser levadas em conta na animação. Sabe-se que os surdos tendem a fazer gestos mais amplos, quando estão conversando com algum ouvinte, e gestos mais curtos quando dialogam entre si. Assim como os ouvintes, que alteram o tom e o volume da voz dependendo do seu estado (alegria, euforia, tristeza, etc.), os surdos também tem os aspectos emocionais refletidos no gesto através da velocidade e da tensão na execução dos mesmos. Dessa forma deve ser analisado velocidade e amplitude do movimento.

A língua oral, assim como a língua falada, varia de país para país. Por exemplo, a língua americana de sinais (ASL) é língua materna dos deficientes auditivos americanos e difere totalmente da língua de sinais britânica (BSL). No Brasil usa-se a língua chamada Libras (Língua Brasileira de Sinais) [de Quadros and Karnopp 2004].

Este trabalho procura determinar um sistema de animação computacional para auxiliar no ensino de gestos em Libras, através da animação mais realista e natural possível, levando em conta um espaço de enunciação dinâmico, como o que ocorre no mundo real. Para que a naturalidade dos gestos seja ainda maior, aspectos como variações na amplitude e velocidade dos gestos, dependendo da distância dos interlocutores, do tom do diálogo e do conhecimento dos mesmos com relação à língua também são investigados.

## 2. Trabalhos Relacionados

Quando se fala em construção de software para pessoas com necessidades especiais, deve-se levar tais necessidades em consideração, o que faz com que ferramentas de auxílio a essas pessoas seja escasso, haja visto que os softwares mais comuns são feitos com foco em pessoas que não possuem tais necessidades. Atualmente já existem alguns softwares para o usuário surdo. Dentre eles alguns já usam animação de humanos virtuais.

O HST (Hand Sign Translator) [Holden and Roy 1992b] [Holden and Roy 1992a] é um tradutor da língua inglesa oral para a língua de sinais chamada *signed english*, criada por educadores australianos. Diferente da língua de sinais australiana (Auslan), *signed english* é uma representação manual do inglês através do uso de sinais e soletramento de palavras através dos dedos. Os sinais usados em *signed english* sempre representam uma letra ou uma palavra e a tradução do inglês é feita por mapeamento direto, pois a sintática e a semântica não se alteram de uma para outra língua. Ele foi criado para ser uma ferramenta educacional e atualmente funciona da seguinte forma: o usuário fornece à aplicação uma entrada textual em inglês e recebe como saída uma animação das mãos gerando a entrada em *signed english*. Por utilizar apenas as mãos é uma língua de mais fácil assimilação, porém não auxilia no aprendizado de Auslan, nem de qualquer outra língua oficial.

O Virtual Signing - ViSiCast [Bangham et al. 2000] é um projeto que procura criar meios de acesso às informações e serviços, em língua de sinais, para cidadãos surdos

---

[Bangham et al. 2000]. Ele consiste no desenvolvimento de sistemas computacionais que utilizam humanos virtuais para animar gestos da BSL (British Sign Language) e foi desenvolvido tendo como base dois trabalhos, o sign-anim [Marshall et al. 2001] e Tessa [Lincoln et al. 1998], que são sistemas tradutores e animadores de texto para língua de sinais. Ambos utilizam captura de imagens através do uso de sensores individuais para as mãos, corpo e face. Porém essa captura possui alguns problemas, tais como a dificuldade de se calibrar os instrumentos, a gravação de uma grande quantidade de sinais, entre outros. Um ponto positivo no ViSiCast é a naturalidade dos gestos, porém o mesmo não trata de detalhes importantes, tais como o espaço de enunciação, as diferentes formas de se executar um gesto (mais tenso, indicando raiva, por exemplo).

Vsigns [Papadogiorgaki et al. 2004] é um gerador de animações a partir da notação de língua de sinais chamada SignWriting [Sutton 1995], muito utilizada pela comunidade surda. Ele gera uma sequência de animações em VRML (*Virtual Reality Modeling Language*). Um ponto positivo dele é a utilização do SignWriting, haja visto que muitos surdos não são alfabetizados por encontrarem dificuldades em assimilar a escrita, porém a animação não é natural e também não trata de questões importantes, tais como o espaço de enunciação.

### 3. Sistema de Animação de Libras

Fazer animações que se baseiam em gestos humanos é uma atividade muito complicada, haja visto que o corpo humano é uma estrutura complexa geralmente representado por um objeto articulado num programa de computador. Este objeto deve possuir várias juntas, com vários graus de liberdade (DOF), onde cada DOF representa um eixo de rotação.

Para que toda essa complexidade de uma aplicação envolvendo movimentação semelhante à dos seres humanos seja tratada, está em desenvolvimento pelo grupo de computação gráfica do Instituto de Informática da UFRGS, o *framework* V-ART (*Virtual Articulations for Virtual Reality* – <http://www.inf.ufrgs.br/cg/v-art>). Este *framework* provê a visualização de objetos 3D, com tratamento de objetos articulados (modelagem e animação).

Para que um sistema gere uma determinada animação de um objeto articulado (um humanóide, por exemplo), deve-se passar por algumas etapas (figura 2): modelagem do objeto, geração das ações, carregamento da cena e geração dos movimentos. A modelagem do objeto envolve a determinação das articulações (ou juntas), juntamente com seus graus de liberdade, ou seja, dado um sistema de coordenadas cartesianas XYZ, definir em quais eixos a junta pode sofrer rotação e quais os seus limites (ângulos mínimo e máximo). O modelo gerado é incluído na cena e usado para gerar as ações a serem animadas. O conjunto de ações é carregado na cena, conforme necessidade.

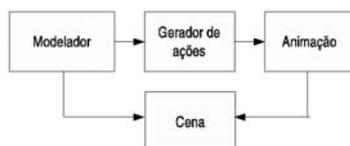


Figure 2. Arquitetura do sistema de animação

## 4. Human Libras

A ferramenta chamada Human Libras, que permite o usuário escolher e visualizar a animação de gestos em Libras foi desenvolvida usando o V-ART. Seguindo a proposta do sistema de animação descrito na seção 3, foi criado um modelo do humanóide e, através de um exportador, gerado um arquivo XML descrevendo a cena. Ele foi utilizado tanto na renderização da cena, como para o gerador de ações.

O gerador de ações (Action Generator) é utilizado para criar os gestos a serem animados e tem como saída um arquivo XML que descreve a animação do gesto. Após a escolha de uma animação existente (gesto em Libras) esta é carregada e executada na cena (figura 3). Um maior detalhamento da ferramenta pode ser visto nas subseções seguintes.

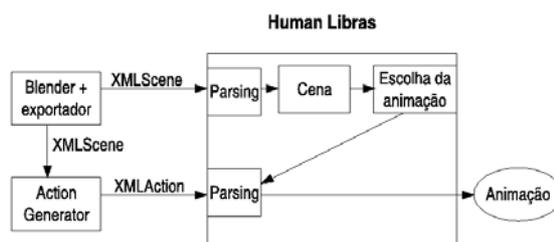


Figure 3. Concepção do Human Libras

### 4.1. Modelagem

O modelo desenvolvido para essa aplicação possui 46 juntas divididas em uniaxiais (um grau de liberdade, ou seja, a junta só se moverá em um eixo), biaxiais (2 graus de liberdade, ou junta se movendo em dois eixos) e poliaxiais (3 graus de liberdade, ou junta se movendo nos 3 eixos), imitando assim as juntas humanas (joelhos como juntas uniaxiais, cotovelos como biaxiais e ombros como poliaxiais, por exemplo) e 46 malhas triangulares, formando as partes do corpo (cabeça, braços, etc.). Ele foi desenvolvido com o auxílio do Blender [Foundation 2006], software para modelagem 3D, disponível sob licença GPL (*GNU General Public License*). Uma imagem do modelo pode ser vista na figura 5. Foi implementado também um exportador, na forma de um *plug-in* para o Blender, que transforma o modelo gerado em um arquivo XML de acordo com os padrões estipulados pelo V-ART.

### 4.2. Criação dos gestos

Para que um movimento seja executado na animação, é necessário especificar o gesto a ser realizado. Entende-se por gesto todo movimento do corpo para exprimir idéias ou sentimentos [Michaelis 1998]. No caso de uma animação, ele indica como as juntas se comportarão com o passar do tempo. Alguns parâmetros estão envolvidos no gesto, tais como as juntas envolvidas no processo, suas posições, a ciclicidade do gesto, sua duração e velocidade.

É necessário que o gesto seja natural, por isso as transições que ocorrem nele devem ser feitas de forma suave, sem quebra de continuidade. A função de interpolação trata dessa suavidade. Ela determina como as rotações das juntas vão se comportar com o passar do movimento. Em busca de uma função que aproxime o gesto virtual do natural,

foram implementadas as funções linear (baseada nas funções lineares, resulta movimentos robóticos, pois a velocidade da execução do movimento das juntas não se altera), *ease-in\_ease-out* (baseada na função seno, gera movimentos mais naturais) e cossenoidal (baseada na função cosseno, não gera movimentos satisfatórios, ou seja, próximos aos movimentos humanos).

Para a criação dos gestos foi desenvolvida uma ferramenta chamada *Action Generator*. Nela o usuário pode mover a junta que escolher de acordo com os limites dos graus de liberdade, além de determinar qual a função de interpolação entre os gestos a ser usada, o tempo de duração do movimento, a velocidade, se há ou não repetição do movimento, entre outros. Este gerador cria então um arquivo XML que determina o gesto obtido em função dos parâmetros relacionados anteriormente. A interface deste gerador pode ser vista na figura 4.

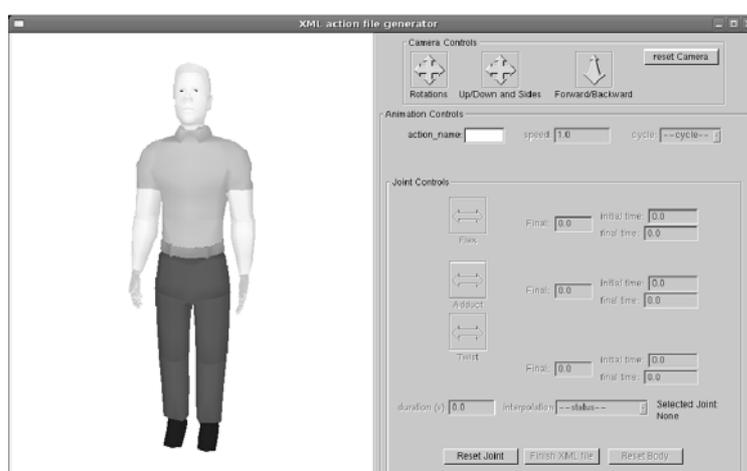


Figure 4. Interface do *Action generator*

### 4.3. Animação

Para que a animação seja gerada é necessário primeiro a renderização da cena, obtida através do modelo criado. Para isso o arquivo XML que descreve o humanoíde é submetido a um *parser* e a cena é então renderizada através de classes específicas do V-ART. Os arquivos XML criados pelo gerador de ações são também submetidos a um *parser*.

Classes do V-ART tratam da manipulação dos dados dos gestos (tempo, duração, juntas envolvidas, ciclicidade, etc) para a geração da animação, que é feita através do V-ART também. Alguns atributos, tais como a velocidade do movimento e sua duração podem ser modificados fora do gerador de ações. Assim, se é desejável que o humanoíde faça uma determinada ação mais rápido (indicando raiva, pressa, etc.) não há necessidade de se gerar novamente a ação e sim apenas modificar esses atributos na aplicação.

### 4.4. Interface

A interface da Human Libras é de fácil utilização. Após carregamento da ferramenta, o usuário deve escolher o gesto que deseja ver animado em Libras. Após essa determinação, basta clicar no botão *Animation* para que se inicie a animação. Se houver desejo de parar uma animação antes de seu término, o botão *stop* deve ser acionado. A animação irá

parar e o humanóide voltará para a posição inicial, considerada posição de descanso. É possível ver o humano virtual em outras posições utilizando-se os controles da câmera, como ilustrado na figura 5.

O espaço de enunciação pode ser exibido ou não, através do acionamento do botão que o habilita. Também é possível diminuir ou aumentar o tamanho desse espaço, permitindo assim a geração da animação do gesto escolhido com base no tamanho total do espaço. Na figura 5 a interface da ferramenta é exibida.

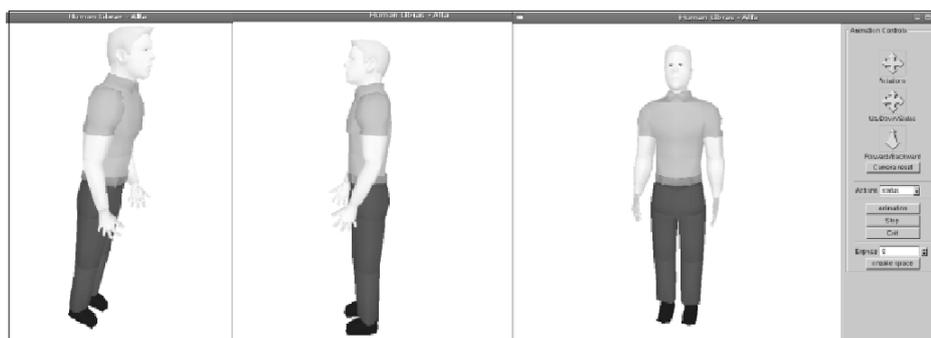


Figure 5. Interface da ferramenta e outras visões do humanóide

## 5. Resultados

A principal contribuição deste trabalho é na determinação de critérios para animação de gestos em Libras, visando auxiliar em seu aprendizado, tanto para surdos quanto para ouvintes. Desta forma, como citado na seção 1, foi necessário analisar alguns critérios, sempre buscando uma maior proximidade do gesto virtual (humanóide) para o real (humano), o que é uma tarefa complicada, haja visto que muitos parâmetros devem ser combinados.

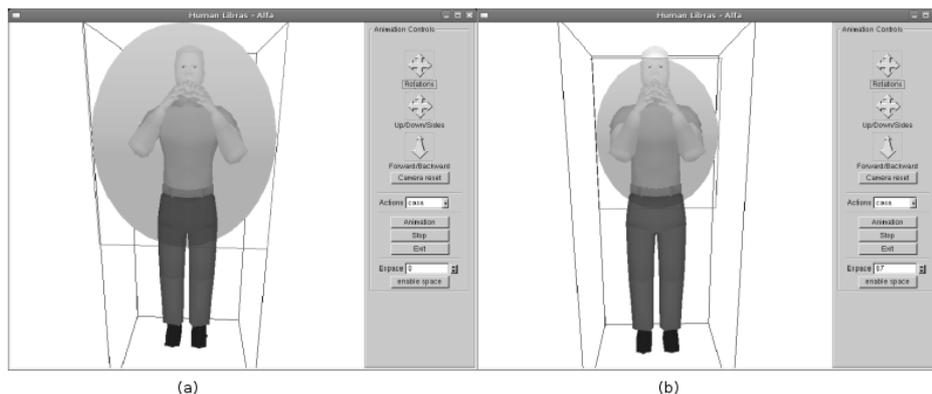
O espaço de enunciação pode sofrer transformações em sua direção ou em seu tamanho. Se houver uma mudança de direção do corpo do humanóide, automaticamente o espaço também sofre essa transformação, já que o espaço faz parte do corpo articulado. Com relação ao tamanho, se houver uma redução do espaço, haverá um rearranjo na posição das juntas, de forma que o gesto fique compreendido dentro do espaço e sem que haja perda de seu significado. Um exemplo pode ser visto nas figura 6.

A naturalidade dos gestos e sua transição, conforme dito anteriormente, são tratados diretamente no interpolador. Dentre os analisados, o *ease-in\_ease-out* apresentou-se como sendo o que gera animações mais próximas às humanas.

Todos os atributos que tratam da temporalidade de Libras podem ser manipulados através do gerador de ações, ou através da aplicação, conforme relatado anteriormente.

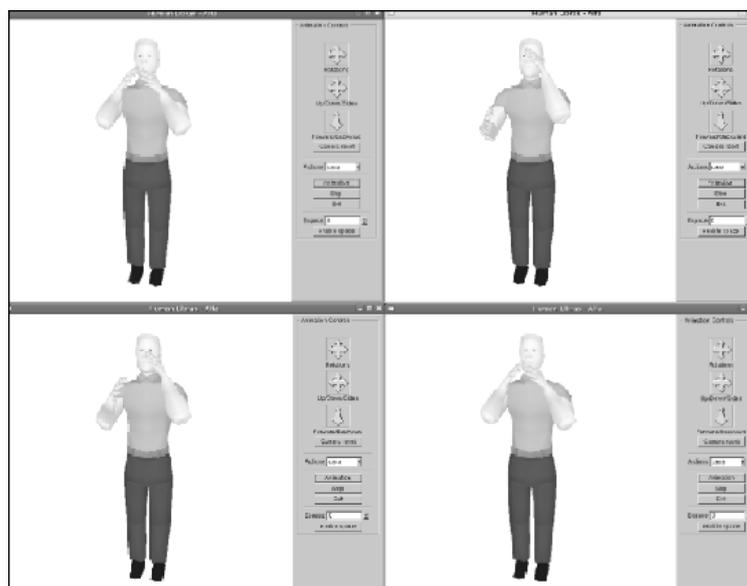
Outra coisa a ser tratada na animação é como combinar dois ou mais gestos de forma harmoniosa. Por exemplo, se o humanóide estiver executando o gesto correspondente à palavra casa e antes de seu término for escolhida a palavra bonito para ser animada, deve ocorrer a interrupção da animação de casa e início da animação de bonito.

Essa combinação é feita através de prioridade nos gestos, ou seja, um gesto pode ter prioridade sobre outro. Assim, se a animação de determinado gesto deve ser executada e este tiver prioridade sobre a que está em execução, as juntas que são comuns aos dois



**Figure 6. Gesticulação da palavra casa com espaço de enunciação padrão (a) e espaço reduzido (b).**

gestos realizarão o movimento do gesto prioritário. Um exemplo pode ser visto na figura 7. Partindo-se do humanóide na posição final da animação da palavra casa, iniciou-se a animação da palavra bonito e, antes de seu término, iniciou-se a animação da palavra casa novamente. Nota-se que na inicialização da palavra casa, a animação de bonito é paralisada.



**Figure 7. Animação de casa antes do término da animação de bonito**

A figura 8 mostra a animação em Libras de duas palavras (casa e bonito). Nota-se que o braço direito do humanóide é usado na primeira animação (casa), mas não é usado na segunda (bonito). Devido a esse fato, as juntas que o compõem voltam à posição considerada de descanso ao mesmo tempo que as juntas participantes da animação atual (no caso a animação da palavra bonito) são modificadas.

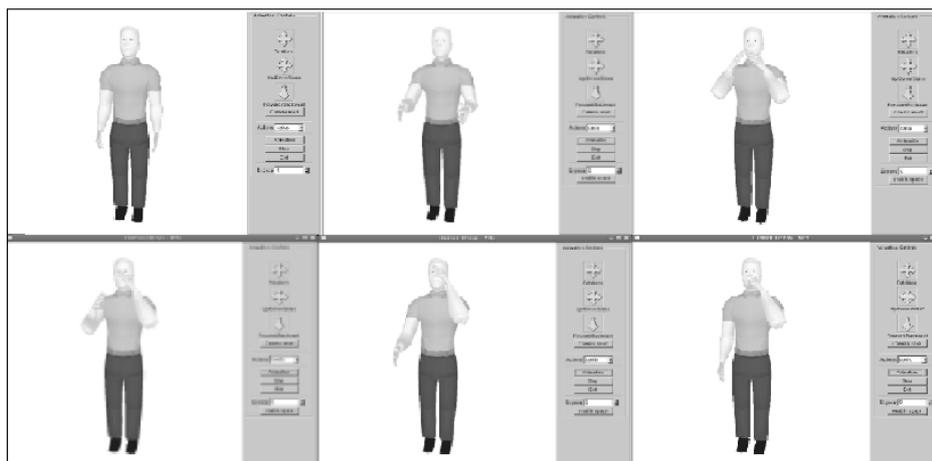


Figure 8. Animação das palavras (primeira linha) casa e bonito (segunda linha)

## 6. Conclusões

O aprendizado de línguas gestuais apoiado por animação computacional ainda é um campo pouco explorado e de grandes possibilidades, haja visto o aumento do uso de animações como auxiliar em diversas áreas do conhecimento.

Foi gerado um sistema de animação de gestos em Libras para auxiliar no ensino da língua, investigando assim parâmetros importantes para a obtenção de uma animação mais realista possível, tais como espaço de enunciação, que é levado em conta na hora da animação dos gestos; velocidade da animação, podendo até indicar mudanças emocionais no humanóide; naturalidade dos gestos, aproximando-o aos gestos humanos; amplitude dos movimentos; interpolação dos gestos, fazendo com que a transição de um para outro seja de forma suave e contínua.

Apesar de ter sido destacada anteriormente a importância das expressões faciais no entendimento dos gestos em Libras, este item não foi analisado neste trabalho, por ser considerado bastante complexo, merecendo portanto um estudo detalhado à parte e sendo considerado um trabalho futuro, na tentativa de se obter gestos mais verossímeis.

Está em desenvolvimento também uma avaliação da ferramenta através de testes com usuários para verificar a veracidade das animações, ou seja, o quanto elas estão naturais. Esse teste será realizado com um grupo de surdos que verificarão o quanto as animações estão próximas dos gestos reais.

## Agradecimentos

As autoras agradecem aos colegas do grupo de computação gráfica da UFRGS que contribuíram no desenvolvimento do *framework V-ART*, a Renato Oliveira pela modelagem do humanóide, à CAPES e ao CNPq por financiarem parcialmente este trabalho.

## References

Bangham, J. A., Cox, S. J., Elliot, R., Glauert, J. R. W., and Marshall, I. (2000). Virtual signing: capture, animation, storage and transmission - an overview of the visicast project. *IEE Seminar on "Speech and language processing for disabled and elderly people*.

- 
- Bideau, B., Multon, F., Kulpa, R., Fradet, L., and Arnaldi, B. (2004). Virtual reality applied to sports: do handball goalkeepers react realistically to simulated synthetic opponents? In *VRCAI '04: Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry*, pages 210–216, New York, NY, USA. ACM Press.
- de Quadros, R. and Karnopp, L. (2004). *Língua de Sinais Brasileira: Estudos Linguísticos*. Artmed, 1st edition.
- Emmelkamp, P., Krijn, M., Hulsbosch, L., de Vries, S., Schuemie, M., and van der Mast, C. (2002). Virtual reality treatment versus exposure in vivo: A comparative evaluation in acrophobia. *Behavior Research & Therapy*, 40(5):509–516.
- Foundation, S. B. (2006). Blender. Página Web. <http://www.blender.org>.
- Holden, E. J. and Roy, G. G. (1992a). The graphical translation of english text into signed english in the hand sign translator system. In Kilgour, A. and Kjeldahl, L., editors, *Computer Graphics Forum (EUROGRAPHICS '92 Proceedings)*, volume 11, pages 357–366.
- Holden, E.-J. and Roy, G. G. (1992b). Learning tool for signed english. In *SAC '92: Proceedings of the 1992 ACM/SIGAPP Symposium on Applied computing*, pages 444–449, New York, NY, USA. ACM Press.
- Karnopp, L. B. (2000). Aquisição da linguagem por crianças surdas. Página Web. [http://www.ines.org.br/ines\\_livros/36/36\\_PRINCIPAL.HTM](http://www.ines.org.br/ines_livros/36/36_PRINCIPAL.HTM).
- Klima and Bellugi, U. (1975). Wit and poetry in american sign language. In *sign language studies*, pages 203–224.
- Lincoln, M., Cox, S. J., and Nakisa, M. (1998). Teh development and evaluation of a speech to sign translation system to assist transactions. *Journal os Human-computer Studies*.
- Marcato, S. A., da Rocha, H. V., and Lima, M. C. M. P. (1974). Um ambiente para a aprendizagem da língua de sinais. In *Sign Language Studies*, number 32.
- Marshal, I., Pezeshkpour, F., Bangham, J. A., Wells, M., and Hughes, R. (2001). On the real time elision of text. *Workshop on Extraction, Filtering and Automatic Summarization*.
- Michaelis (1998). *Michaelis: moderno dicionário da língua portuguesa*. Companhia Melhoramentos, São Paulo, Brasil.
- Papadogiorgaki, M., Grammalidis, N., Sarris, N., and Strintzis, M. G. (2004). Synthesis of virtual reality animations from swml using mpeg-4 body animation parameters. In *Workshop on the Representation and Processing of Sign Languages, 4th International Conference on Language Resources and Evaluation LREC 2004*.
- Sutton, V. (1995). Signwriter manual. In *Deaf Action Committee for SignWriting*.