

# Processo de Desenvolvimento de Interfaces de Usuário Baseadas em Gestos

Anna C. S. Medeiros  
Universidade Federal da Paraíba  
Centro de Informática  
João Pessoa – PB - Brasil  
linnamedeiros@gmail.com

Tatiana A. Tavares  
Universidade Federal da Paraíba  
Laboratório de Aplicações de Vídeo  
Digital (LAVID)  
Centro de Informática  
João Pessoa – PB - Brasil  
tatiana@lavid.ufpb.br

Iguatemi E. da Fonseca  
Universidade Federal da Paraíba  
Centro de Informática  
João Pessoa – PB - Brasil  
iguatemi@ci.ufpb.br

## RESUMO

A utilização da nossa linguagem corporal para se comunicar com sistemas computacionais é uma característica cada vez mais possível e aplicável no mundo real. Esse fato é potencializado pela evolução das soluções comerciais baseadas em reconhecimentos de gestos. Uma interface de gestos complementa ou substitui a forma de navegação de uma interface convencional, cabe ao desenvolvedor escolher a opção mais adequada para a sua aplicação. Assim quando optamos pelo uso de gestos, os gestos serão responsáveis por acionar as funções oferecidas pelos sistemas. O presente trabalho apresenta um processo de desenvolvimento de interfaces de usuário centrado na escolha de gestos. O processo proposto deve auxiliar designers de interfaces a integrarem interação natural por gestos em suas aplicações de forma mais sistemática.

## ABSTRACT

The use of our body language to communicate with computer systems is an increasingly possible and applicable feature in the real world. This fact is potentiated by the evolution of gesture recognition based commercial solutions. A gesture interface complements or replaces navigation in a conventional interface, it is up to each developer to choose the most appropriate option for their application. Therefore when opting for gesture usage, the gestures will be responsible to activate the systems functions. This work presents an interface development process centered on the selection of gestures. The proposed process should help interface designers to incorporate gesture based natural interaction into their applications in a more systematic way.

## Keywords

Gestures, Natural Interaction, Natural User Interface.

## 1. INTRODUÇÃO

A partir do estudo e pesquisa acerca das formas de interação entre o homem e os dispositivos eletrônicos, usando como meio seu próprio corpo, foco de estudo da Interação Natural (IN), verificou-se que existem diversas possibilidades de interação: através de comandos de voz, dispositivos hápticos, olfato, locomoção, gestos ou detecção e identificação de partes do corpo humano como rosto, mão, polegar, retina, entre outros [1].

É importante destacar que a Interação Humano-Computador (IHC) se dedica aos fenômenos de comunicação entre pessoas e sistemas computacionais, encontrando-se na interseção de três ciências: comportamentais; sociais; computação e informação, envolvendo todos os aspectos relacionados com a interação entre usuários e sistemas [2]. A pesquisa sobre o estado da arte da interação humano-computador despertou para investigar como

usar a interação natural como uma alternativa a interface convencional, podendo tornar mais simples e intuitiva a interação com sistemas computacionais a partir da inserção de IN no ambiente.

É sabido que a população em geral lida com tecnologia diariamente, e que existem aplicações para os mais diversos públicos-alvo, assim, a simplicidade torna-se necessária para facilitar a interação com sistemas e leva a um relacionamento mais fácil e sustentável com a mídia e tecnologia [3].

O uso de tecnologias no sentido de prover eficiência para interações naturais, a partir de gestos mais próximos dos usados no cotidiano, mais intuitivos, tem o potencial para simplificar a Interação Humano-Computador. Com o avanço tecnológico e a criação de novos dispositivos para interação natural, dispositivos atuais tiveram o seu custo reduzido rapidamente, o que passou a viabilizar o uso destes em mais aplicações, abrangendo mais usuários. Isso ocorre também pelo uso destas tecnologias na indústria de entretenimento, gerando dispositivos mais acessíveis, como é o caso do Microsoft Kinect, utilizado inicialmente para interagir via gestos com o console de jogos eletrônicos Xbox 360.

Considerando que um vocabulário de gestos pode ser entendido como uma ferramenta importante para a interação humano-computador e que a popularização dos dispositivos tecnológicos os deixa mais acessíveis, gerando um aumento na diversidade de usuários, destaca-se um espaço para pesquisas nessa área, para definir uma melhor interação para pessoas de diferentes culturas, necessidades e formação, pois gestos são muito dependentes do contexto cultural em que o indivíduo está inserido [4].

Existem trabalhos que focam na tecnologia perceptiva e no desenvolvimento do software, desconsiderando elementos referentes à determinação dos gestos, um exemplo é [5] que, apesar de ser de grande importância para a área, não procura o entendimento dos gestos em busca da escolha mais adequada para as interfaces criadas, o que pode levar a utilizar gestos cansativos, pouco intuitivos ou não funcionais, podendo prejudicar o desempenho da aplicação.

O processo de associar um gesto à determinada ação ou função do sistema não é trivial, pois deve levar em conta uma série de fatores tais como ergonomia, intuitividade e objetividade [6][7], o que destaca a importância da presente pesquisa.

A pesquisa aqui descrita trabalha na construção de um processo que possa auxiliar na decisão sobre que gesto representaria uma determinada função de um sistema, para tanto foi realizado um levantamento de trabalhos que foquem no processo de desenvolvimento de gestos, comparando-os e assim produzindo

um processo que englobe os melhores aspectos dos trabalhos analisados.

## 2. TRABALHOS CORRELATOS

Para melhor compreensão dos trabalhos correlatos, os seguintes parâmetros de comparação foram estabelecidos:

- **Modelo de Processo de Desenvolvimento:** trabalhos que indicam modelos ou processos que foram utilizados para o desenvolvimento de interfaces em IN;
- **Vocabulário:** considerou-se que um trabalho aborda o vocabulário se ele identifica qualquer conjunto, banco ou base de gestos. Esse vocabulário deve apresentar uma descrição do gesto que pode ser textual e não precisa identificar a funcionalidade do mesmo;
- **Templates:** trabalhos que abordam templates de gestos implicam em ter uma representação gráfica do gesto, tornando-o fácil de ser entendido e utilizado em qualquer tipo de aplicação;
- **Aplicações:** trabalhos que abordam aplicações mostram exemplos (cases) que utilizam gestos como forma de interação. Essas aplicações podem ser de qualquer tipo, por exemplo, baseadas em PCs ou dispositivos móveis;
- **Avaliação com usuários:** considera-se que um trabalho relaciona avaliação com usuários quando este testa de alguma forma o gesto desenvolvido com usuários, obtendo assim um *feedback* sobre o uso do mesmo.

### 2.1 Detalhamento dos Trabalhos Correlatos

O trabalho descrito em [6] e [8] apresenta dois processos para desenvolver gestos para interfaces gestuais, um é voltado para a tecnologia e o outro é voltado para o lado humano. Esse mesmo processo também é descrito posteriormente em [7]. O processo voltado para tecnologia se preocupa em utilizar gestos que sejam facilmente reconhecidos pelos artefatos tecnológicos da época, porém constatou-se que os usuários taxaram os gestos resultantes desse processo como estressantes, ilógicos e até impossíveis. O processo voltado para o lado humano favorece gestos fáceis de lembrar e executar pelo usuário, intuitivos, lógicos, funcionais e ergonômicos. Esse procedimento é dividido em quatro etapas, A, B, C e D.

Na etapa A devem-se definir todas as possíveis funções/ações que serão acionadas por meio de gestos. Na etapa B, devem-se definir gestos protótipos que irão corresponder a cada função, isso deve ser feito através de entrevistas com voluntários, observando-os através de cenários de onde comunicarão para um avaliador as mesmas mensagens que iriam comunicar ao computador. Devem-se gravar as entrevistas para análise posterior. Segundo Nielsen et al., é importante afastar pensamentos técnicos nessa etapa, quanto menos o voluntário souber dos aspectos técnicos da ferramenta, melhor. O número de pessoas necessárias para esta investigação depende de quão amplo o grupo de utilizadores é e quão diverso os resultados do teste são.

Na etapa C, o procedimento analisa nas gravações das entrevistas da etapa B os gestos mais frequentes, a ergonomia dos mesmos, força interna causada por cada gesto, desvio da posição neutra, angulações extremas, forças aplicadas nas articulações e duração dos gestos.

Por último a etapa D, na qual o procedimento avalia o vocabulário de gestos através de três testes: teste de atribuição de semântica, de memória e de stress. Cada teste possui uma pontuação para poder medir a qualidade do resultado final do vocabulário de

gestos, podendo até mesmo optar por repetir o procedimento descrito por Nielsen et al..

No teste de semântica, são apresentados ao voluntário os gestos escolhidos na etapa anterior e a lista de funções, porém não são reveladas as correspondências entre gesto e função, fica a cargo do voluntário intuir as funções relativas aos gestos. A pontuação é igual à soma dos palpites errados dividido pelo número de gestos.

Os testes de memória medem o conforto e familiaridade do usuário com os gestos, só depois de o usuário poder usar os gestos confortavelmente é que ele vai manter o foco na aplicação em si e não na interface. Primeiro é mostrado o vocabulário de gestos para cada voluntário, depois um slide com as funções, sendo 2 segundos por função. O usuário deve então fazer o gesto correspondente a cada função até obter todas corretas; caso erre, o procedimento do slide é reiniciado. O vocabulário de gestos é revisado entre cada reinicialização do slide e aqui a pontuação corresponde ao número de reinicializações. O resultado é um indicador de quão difícil será para um novo usuário se tornar confortável com o vocabulário de gestos, para que estes não roubem o foco das tarefas do usuário.

O teste de stress apresenta ao voluntário uma lista sequencial de gestos, cada voluntário deve repetir a sequência uma determinada quantidade de vezes, onde tal quantidade é determinada pelo resultado da divisão de duzentos pelo tamanho do vocabulário de gestos. Depois de concluído o teste, o voluntário é questionado sobre quão estressante era cada gesto, fornecendo uma classificação geral para cada gesto usado durante um longo tempo, por exemplo, “sem problema”, “ligeiramente cansativo”, “irritante”, “doloroso” ou “impossível”.

Para testar o procedimento voltado para o lado humano, é usada uma aplicação hipotética que permite que o usuário coloque objetos virtuais 3D em cima de uma mesa, mova-os e altere as configurações de estilo sobre eles. As funções determinadas para essa aplicação foram “Ativar Menu”, “Selecionar”, “Selecionar Todos”, “Inserir”, “Mover”, “Escala”, “Rotacionar”, “Deletar”, “Sim/Confirmar”, “Não/Desfazer” e “Copiar-Colar”.

Três cenários foram construídos na etapa B para essa aplicação: o primeiro (Figura 1) é um bar onde o *bartender* não consegue escutar o cliente devido ao excesso de barulho no ambiente fictício, o cliente deve então proceder com gestos para obter seus pedidos; o segundo cenário consiste em um conversar sobre os possíveis posicionamentos de mobília, janelas e portas em um ambiente desenhado no papel e esse cenário serve como auxiliar do cenário seguinte; no terceiro cenário (Figura 2), o voluntário, agora já familiarizado com a cena, deve comunicar onde quer colocar os objetos da cena (mobília, porta e janelas) usando gestos. Todos os cenários foram gravados.

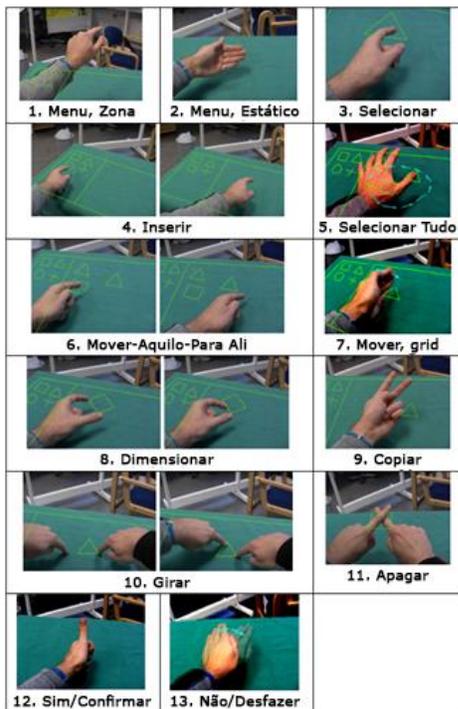


Figura 1 - Primeiro cenário



**Figura 2 - Terceiro cenário**

No passo C, são analisados os vídeos da etapa anterior e são escolhidos os gestos finais (Figura 3) baseados nos parâmetros anteriormente mencionados.



**Figura 3 - Gestos finais da aplicação com objetos virtuais 3D**

Na etapa D, são realizados testes com o vocabulário de gestos apresentado na Figura 3. Os resultados dos testes são demonstrados na Figura 4.

	Média	Min Max	Divergência
Teste 1 - Semântica	0.10	0.0-0.33	0.02
Teste 2 - Memória	2.13	0.0-6.0	3.55
Teste 3 - Stress	2.5	1.0-3.0	0.57

**Figura 4 - Resultado dos testes do passo D**

Segundo Nielsen *et al.*, a análise comparativa de um único vocabulário de gesto requer que seja estabelecida uma meta para se atingir com os resultados, o que pode ser difícil de predefinir. O melhor seria comparar diferentes vocabulários de gestos.

A pesquisa de Henze *et al.* é baseada no trabalho de Nielsen *et al.*. Henze *et al.* fazem algumas modificações no processo descrito por Nielsen *et al.* e aplica-as no desenvolvimento de um software para tocar músicas, que não foi implementado. [9]

O processo descrito por Henze *et al.* também consiste em quatro etapas: contexto de uso e funções, design participatório, definição de conjunto de gestos e a última etapa, chamada avaliação e melhoramentos.

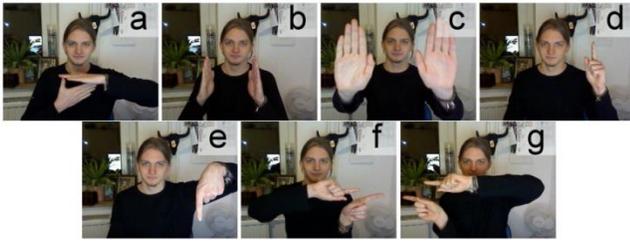
Na etapa de contexto de uso e funções, Henze *et al.* procurou entender o contexto de uso de aplicações que tocam música no intuito de começar a definir as funções que seu tocador de músicas hipotético teria. Esse estudo foi dividido em duas partes: na primeira parte o foco foi a análise da necessidade e objetivo de escutar música em diferentes situações, através de entrevistas com voluntários; na segunda parte foi apresentado aos voluntários um protótipo que reconhece capas de CD's via webcam e toca uma música do álbum apresentado, com o intuito de obter *feedback* inicial sobre a funcionalidade e o comportamento desejado do sistema hipotético. Ao final desta etapa, obtêm-se um conjunto inicial de funcionalidades para a aplicação hipotética. Nove pessoas participaram dessa etapa.

O objetivo da etapa de design participatório é validar o conjunto das funções definidas na etapa anterior e derivar múltiplos gestos, para cada função, através de técnicas de design participatório. Esse estudo também foi dividido em duas partes. Na primeira parte foi validado o conjunto de funções anterior e foram coletadas mais informações a respeito dos hábitos de escutar música dos participantes voluntários, através de entrevistas pessoais. O conjunto de funções finais foi: “play”, “stop”, “pause”, “próximo”, “anterior” e “modificar volume”. Na segunda parte o foco foi coletar potenciais gestos. Para isso, os voluntários executavam gestos para cada função presente em uma lista de funções dada, na frente de um computador equipado com uma câmera (Figura 5). Os participantes deveriam anunciar oralmente que função que desejavam executar antes de gesticular. A funcionalidade do sistema era simulada por um operador de acordo com a respectiva função anunciada pelo voluntário participante. Ao final desta etapa, foram obtidas gravações com os gestos candidatos e eles foram classificados como: estáticos, dinâmicos ou híbridos. Dez novos voluntários participaram dessa etapa.

Na etapa de definição de conjunto de gestos, o objetivo é formalizá-los e definir seus conjuntos consistentes. Usando as gravações da etapa anterior pôde-se analisar os gestos escolhidos levando-se em consideração, além de conceitos também considerados na etapa C de Nielsen *et al.*, a popularidade ou frequência dos gestos candidatos. Ao final dessa etapa, diferente do passo B do estudo de Nielsen *et al.* que define um gesto protótipo para cada função, o trabalho de Henze *et al.* define dois conjuntos de gestos protótipos, um estático (Figura 6) e outro dinâmico (Figura 7) para cada função, para serem analisados nas próximas etapas.



**Figura 5 - Exemplo de imagem da câmera do estudo de design participatório**

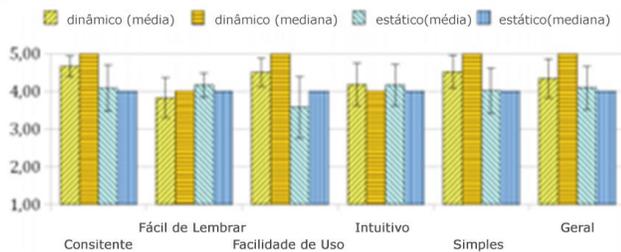


**Figura 6 - Gestos Estáticos; (a)Play, (b)Pause, (c) Stop, (d) Aumentar Volume, (e) Diminuir Volume, (f) Próximo, (g) Anterior**



**Figura 7 - Gestos Dinâmicos; (a)Play/Pause, (b)Stop, (c) Aumentar volume, (d) Próximo**

Na etapa de avaliação e melhoramentos, os dois conjuntos de gestos protótipos são avaliados e eventualmente refinados. Além disso, é feita outra pesquisa a respeito das funções, se existem outras além das propostas que seriam essenciais para o controle de um tocador de músicas. Essa etapa é dividida em três partes. A primeira parte avalia o conjunto de funções já determinado e a oportunidade para o participante comentar sobre novas funções. Foi usada uma escala comparativa de 1 (desnecessário) até 7 (essencial) para avaliar o conjunto de funções e todas as funções foram classificadas com uma média e mediana de, pelo menos, 5,5. A segunda parte consiste num teste com usuário usando um ambiente similar ao descrito na etapa de design participatório, no qual cada voluntário participante deveria executar os gestos estáticos e dinâmicos de cada função apresentada através de uma lista de funções. Após o término da avaliação, cada participante classificava a simplicidade, intuitividade, consistência, facilidade de uso e facilidade de lembrança dos gestos (Figura 8). A escala ia de 1 (nem um pouco) até 7 (perfeito) e, na última parte, o participante poderia dar a sua opinião sobre alguma função que ele sentisse falta. Duas pessoas não mencionaram funções extras, outras citaram funções como “ir para música aleatória” e “editar lista de músicas”, dando exemplos de gestos para elas. Doze voluntários participaram dessa etapa. A etapa final do trabalho de Henze *et al.* não utiliza os testes de semântica, memória e stress, que são descritos na etapa D de Nielsen *et al.*



**Figura 8 - Classificação dos gestos estáticos e dinâmicos, usando médias e medianas**

O trabalho descrito em [10] desenvolve gestos para funções de sistemas horizontais humano-computador, como *tabletops*, para investigar as preferências gestuais dos usuários desses sistemas.

Os autores definiram 36 funções, sendo as principais: “selecionar”, “abrir”, “mover”, “scroll”, “desenhar”, “cortar/colar”, “rotacionar” e “zoom”. Elas foram realizadas por voluntários em um sistema horizontal humano-computador. Para cada função, a tela exibia uma imagem estática e era pedido ao usuário para usar gestos na realização hipotética das funções, utilizando os objetos exibidos na imagem. Vinte voluntários participaram do estudo, sendo treze homens e sete mulheres. A frequência foi usada com critério de seleção dos gestos.

O trabalho de Choi *et al.* foca em mostrar que os gestos mais populares (mais frequentes) em uma pesquisa com usuário não são necessariamente a melhor opção. Para validar suas hipóteses, ele usa como estudo de caso um sistema *smart-home* [11] embasando seu estudo na análise dos trabalhos: [6], [8], [9], [10], entre outros. Estas propostas apresentam processos de desenvolvimento voltados para gestos nos quais a frequência é utilizada como parâmetro de escolha. Cada candidato pensa em um universo restrito de possibilidades de gestos para cada função, se um dos candidatos for um designer de interfaces ou mesmo um usuário mais experiente então suas sugestões podem ser valiosas, sendo assim é provável que gestos significativos sejam negligenciados devido à sua baixa frequência.

Os autores criaram hipóteses que caracterizavam as falhas que os mesmos perceberam e conduziram então dois experimentos para identificar as falhas supracitadas. O primeiro experimento (Figura 9) focou no mapeamento de gestos para a interação com estruturas de uma casa e eletrodomésticos e estudou as razões das escolhas de gestos de cada participante. No segundo experimento (Figura 9), os participantes foram convidados a escolher o gesto mais adequado para cada comando, a partir de uma lista que incluiu a maioria dos provenientes do primeiro experimento. Os comandos foram definidos para o cenário de uma *smart-home*. No primeiro experimento, houve 30 participantes e, no segundo, 28. Todos os participantes do segundo experimento estavam presentes no primeiro. Após ambos os experimentos, cada participante avaliava a adequabilidade do gesto escolhido.



**Figura 9 - Representação visual dos dois primeiros experimentos. Fonte: [11]**

Durante um terceiro experimento, foram analisados e comparados os resultados de ambos os experimentos e então se constatou que a maioria dos participantes mudou sua escolha de gesto no segundo experimento e alguns dos gestos, que apenas alguns participantes provenientes do primeiro experimento realizaram, foram selecionados como os melhores durante o segundo experimento, mostrando assim que gestos com poucas ocorrências podem ser ideais.

Choi *et al.* não concluem qual seria o gesto ideal para as ações escolhidas dentro do cenário de uma *smart-home*, alegando que o foco do trabalho não era esse.

Em [12], Bodiroza *et al.* descreve o processo utilizado para desenvolver dois tipos de vocabulário de gestos, um humano (HGV) e um robótico (RGV), para uma interface humano-robô em um cenário de Robô-garçon.

A metodologia para o HGV consiste em pedir para voluntários criarem um gesto para certa função dada. Caso não haja um percentual de gestos similares acima de 25% para uma determinada função, deve-se então repetir o procedimento para essa função, só que agora dando opções de gestos para os voluntários.

A metodologia para o RGV consiste em mostrar vídeos de robôs realizando gestos e perguntar aos voluntários se tais gestos correspondem à função desejada. No referido artigo, somente a metodologia para o HGV foi executada.

O trabalho [13] gerou um vocabulário de movimentos contendo 4.080 movimentos humanos a partir de gravações, usando um sistema de câmera de alta velocidade de captura. A partir do uso do 3D Studio MAX e MATLAB, os autores processaram e converteram os dados de captura de movimento em formatos úteis para animação.

O vocabulário contém dados de 30 indivíduos e cada indivíduo disponibilizou os dados relativos às seguintes ações consideradas, pelos autores, comuns no cotidiano: andar, bater (em uma porta), jogar (uma bola) e levantar (uma lata de spray), bem como para as sequências dessas diferentes ações. Para todas as ações, os movimentos foram capturados em diferentes emoções (raiva, felicidade, neutralidade e tristeza) e para as ações de bater, jogar e levantar foram obtidas dez repetições de cada movimento em cada emoção.

Segundo os autores, o uso em potencial desse vocabulário inclui a análise computacional e de comportamento de como gênero, emoção e identidade são codificados e decodificados do movimento humano. Também auxilia na síntese de animação de caráter humano, já que tal vocabulário provê uma coleção de movimentos de ação única, como também sequências desses movimentos. Os autores ainda mencionam que tal vocabulário também seria adequado para o estudo das fontes de variabilidade nos movimentos, já que um mesmo ator realiza repetidas vezes cada movimento.

A pesquisa de Bon-Woo Hwang [14] trata da construção de um banco de dados de gestos corporais, contendo 14 gestos normais do cotidiano, realizados por 20 participantes com idade entre 60 e 80 anos; 10 gestos anormais, caracterizando situações de emergência como uma queda repentina; e 30 gestos de comando, tais como apontar, realizados por 20 participantes com idades entre 10 e 30 anos. O banco de dados é formado por dados de movimentação 3D, vídeos estéreo 2D e dados de silhueta 2D.

Gestos normais : sentar numa cadeira, levantar de uma cadeira, andar até um local, tocar o joelho e a cintura, levantar a mão direita, esticar uma mão, flexionar a cintura, sentar no chão, abaixar-se até o chão, deitar no chão, acenar uma mão, andar para frente e andar circularmente.

Gestos anormais : estava em pé e caiu para frente, estava em pé e caiu para trás, estava em pé e caiu para esquerda, estava em pé e caiu para a direita, estava sentado no chão e caiu para a esquerda, estava sentado no chão e caiu para a direita, estava sentado no chão e caiu para trás, estava sentado numa cadeira e caiu para a esquerda, estava sentado numa cadeira e caiu para a direita, estava sentado numa cadeira e caiu para frente.

Gestos de comando : sim, não, mover a mão para frente, mover a mão para trás, mover a mão para cima, mover a mão para baixo, mover a mão para a direita, mover a mão para a esquerda, apontar abaixo-esquerda, apontar abaixo-centro, apontar abaixo-direita, apontar meio-esquerda, apontar meio-centro, apontar meio-direita, apontar cima-esquerda, apontar cima-centro, apontar cima-direita,

desenhar “O”, desenhar “X”, sinalizar para chamar alguém se aproximar, desenhar ‘0’, desenhar ‘1’, desenhar ‘2’, desenhar ‘3’, desenhar ‘4’, desenhar ‘5’, desenhar ‘6’, desenhar ‘7’, desenhar ‘8’, desenhar ‘9’.

Segundo os autores, o banco de dados pode ser potencialmente usado no desenvolvimento de algoritmos de reconhecimento de gestos 3D que usem informações de profundidade a partir de câmeras estéreo, na avaliação de algoritmos estimativos de componentes corporais 2D/3D a partir de cores ou imagens do tipo silhueta de vários gestos, na avaliação de algoritmos de reconhecimento de gestos 2D sob variações de pose, na avaliação de algoritmos de reconhecimento de gestos a partir da posição ou ângulo de componentes do corpo.

O trabalho [15] trata de uma técnica para anotação de nuvem de pontos baseada na entrada de gestos a mão livre, chamado Slice-n-Swipe (Figura 10). Essa técnica usa refinamento progressivo para possibilitar que o usuário especifique os pontos de interesse. O Slice-n-Swipe usa o LeapMotion para permitir aos usuários cortar interativamente o *dataset* e assim remover os pontos indesejados.



Figura 10 - Slice-n-Swipe em ação

Essa ferramenta trabalha com 2 gestos: cortar (*slice*) e roubar (*swipe*), sendo que o *swipe* tem de ser sempre perpendicular ao gesto de cortar. Além disso, o usuário pode usar o teclado para adicionar comentários e controla a câmera com a mão não dominante através de um mouse especial (3DConnexion SpacePilot Pro 3D Mouse). Tal trabalho é apenas um exemplo de aplicação que emprega gestos, mas não utiliza nenhum processo de desenvolvimento de vocabulário de gestos, *template*, nem mesmo testa a aplicação final em relação à satisfação dos usuários a respeito da naturalidade da interface.

Em [16], é proposto um sistema de controle específico para TV que possui câmera única. Tal sistema explora uma alternativa ao controle remoto nas funções mais frequentemente utilizadas ao assistir TV. Para determinar essas funções, utilizou-se um questionário aberto ao público, assim obtiveram-se como mais populares as seguintes funções: volume, canal, stop/play e busca rápida.

Não houve um processo específico para definir os gestos correspondentes às funções mencionadas que foram determinados pelos próprios desenvolvedores. O *template* dos gestos escolhidos era visível na tela da TV (Figura 11).



Figura 11 - Ambiente do trabalho desenvolvido. Fonte: [16]

Após o desenvolvimento do protótipo do trabalho, 15 voluntários utilizaram-no e responderam um questionário no qual a maioria dos participantes concordou que o sistema proposto pode ser usado de forma conveniente e divertida ao se familiarizarem com ele, porém houve reclamações sobre o conforto e diversidade dos gestos, o que ressalta a importância de um bom processo de desenvolvimento de gestos para melhorar a experiência final do usuário.

## 2.2 Análise Comparativa

Na Tabela 1 podem-se comparar os trabalhos citados anteriormente. Dentre eles pôde-se observar que existem trabalhos sobre processo de desenvolvimento de gestos que não se preocupam em aplicar o vocabulário final no desenvolvimento de um protótipo. Outros trabalhos apenas desenvolvem um vocabulário de gestos sem se preocupar com o processo de desenvolvimento dele. Alguns trabalhos não esboçam um *template* dos seus gestos finais, o que pode dificultar o entendimento ou mesmo a repetição de um experimento realizado neles. Pode-se ainda observar trabalhos que se preocupam em avaliar a experiência do usuário final.

Tabela 1 - Comparação entre os trabalhos anteriormente descritos

Trabalho	Processo de Desenvolvimento	Vocabulário	Template	Aplicação	Avaliação com Usuários
[6]	X	X			
[9]	X	X			
[10]	X	X			
[11]	X	X	X		
[12]	X	X			
[13]		X			
[14]		X	X		
[15]		X		X	
[16]		X	X	X	X

## 3. PROPOSTA

O presente trabalho formula um processo de desenvolvimento de gestos para funções/ações de uma dada aplicação, no intuito de auxiliar designers de interfaces na utilização de gestos em suas aplicações. Ao final desse processo, obtêm-se um vocabulário e *template* de gestos dedicados á aplicação em questão, prontos para a integração.

O processo de desenvolvimento de gestos para funções/ações de uma dada aplicação possui cinco etapas (Figura 12). Na primeira etapa é analisado o domínio do problema e são determinadas as

funções da aplicação a ser desenvolvida, por exemplo, rotacionar, aproximar/afastar, selecionar. Essa etapa não se preocupa com nenhum tipo de gestos, apenas em determinar as ações/funções às quais se deseja atribuir gestos.

A segunda etapa necessita de testes com voluntários para gerar gestos-protótipos para as de funções determinadas na etapa anterior. É criado um ou mais cenários de teste que abstraíram pensamentos técnicos e ao mesmo tempo contextualizam as funções desejadas para estimular o voluntário a executá-la, todas as interações nessa etapa são gravadas. O número de participantes deve ser diretamente proporcional à diversidade dos gestos obtidos para cada função, ou seja, à medida que os testes são aplicados, quanto mais diversos os gestos atribuídos a uma dada função, mais voluntários serão necessários.

Na terceira etapa são analisadas as gravações da etapa anterior e são selecionados os gestos que participarão da próxima etapa, levando em consideração os mais frequentes, os escolhidos pelo designer e aqueles cuja movimentação e postura não comprometam estruturas físicas como articulações. É interessante gerar um *template* dos gestos escolhidos, para ser utilizado nas próximas etapas.

A quarta etapa pode ser executada de duas formas: dispo de novos voluntários deve-se executar o(s) cenário(s) da segunda etapa, exceto que agora com uma lista de sugestões de gestos, selecionados na terceira etapa, podendo o voluntário escolher algum da lista ou criar um novo para dada função. A outra forma seria feita uma entrevista explicando as outras opções de gestos escolhidas na etapa anterior e perguntando se o voluntário mudaria sua escolha de gesto. Ao obter os novos resultados das escolhas dos participantes desta etapa, são definidos “N” gestos finalistas por função, totalizando “N” vocabulários de gestos para a lista de funções, para serem usados na etapa posterior, onde “N” pode variar de um a infinito, cabe ao designer definir, no decorrer dos testes, dependendo de quão diversos forem os gestos fornecidos pelos voluntários para cada função.

Na quinta e última etapa, o processo avalia os “N” vocabulários de gestos através de três testes: teste de atribuição de semântica, de memória, e de stress, conforme descrito por Nielsen *et al.*. Cada teste tem sua pontuação e no final será possível decidir o melhor vocabulário através da comparação de suas pontuações. No caso de ser apenas um vocabulário, a pontuação servirá como base para inferir a qualidade dele.

Nos testes que seguem, os “N” vocabulários de gestos obtidos serão testados separadamente, os mesmos voluntários irão avaliar ambos os vocabulários, variando a ordem de avaliação dos vocabulários para cada voluntário, ou seja, no caso de serem dois vocabulários, um voluntário começará avaliando o vocabulário A e posteriormente o B, já outro voluntário começaria pelo B, de modo que no final a quantidade de avaliações começando pelo A seja igual ou próxima da quantidade começando pelo B.

O teste de atribuição semântica consiste em apresentar ao voluntário um vocabulário de gesto e a lista de funções, porém não são reveladas as correspondências entre gesto e função, fica a cargo do voluntário intuir as funções relativas aos gestos. A pontuação é igual à soma dos palpites errados, dividido pelo número de gestos.

Os testes de memória medem o conforto e familiaridade do usuário com os gestos e só depois de o usuário poder usar os gestos confortavelmente é que ele vai manter o foco na aplicação

em si e não na interface. Primeiro é mostrado um vocabulário de gestos para cada voluntário, depois um slide com as funções, sendo 2 segundos por função. O voluntário deve então fazer o gesto correspondente a cada função até obter todas corretas, caso erre, o procedimento do slide é reiniciado. O vocabulário de gestos é revisado entre cada reinicialização do slide, aqui a pontuação corresponde ao número de reinicializações. O resultado é um indicador de quão difícil será para um novo usuário se tornar confortável com o vocabulário de gestos, para que os gestos não roubem o foco das tarefas do usuário.

O teste de stress apresenta ao voluntário uma lista sequencial de gestos. Cada voluntário deve repetir a sequência uma determinada quantidade de vezes que é definida pelo designer e deve ser suficiente para o voluntário poder inferir sobre qualquer possível desconforto. Depois de concluído o teste, o voluntário é questionado sobre quão estressante é cada gesto, fornecendo uma classificação geral para aquele usado durante um longo tempo, por exemplo, “sem problema”, “ligeiramente cansativo”, “irritante”, “doloroso” ou “impossível”.

Ao final dos testes, podem-se comparar os vocabulários através de suas pontuações e classificações, então seria desenhado um *template* final para os gestos do vocabulário escolhido, assim poder-se-ia dar início à implementação desses para suas respectivas funções na aplicação escolhida.

Para validar o processo supracitado, realizamos um piloto no contexto de um jogo que faz uso da Interação Natural. O jogo Fruits Ninja (Figura 13) foi criado em 2010 pela empresa HALFBRICK Studios e consiste em basicamente cortar rapidamente variados tipos de frutas que vão aparecendo em diferentes pontos da tela, contendo apenas uma função: a função “cortar”.



**Figura 12 - Representação do Processo de Definição de Gestos para Funções**



**Figura 13 - Fruit Ninja**

O experimento piloto foi realizado no laboratório LAVID, da UFPA, e contou com a participação de 10 voluntários no período de 1 a 7 de junho de 2014. Os voluntários foram contextualizados sobre o objetivo da pesquisa e apresentados ao cenário de testes da segunda etapa, pois a primeira consiste em determinar as funções da aplicação a ser desenvolvida e no estudo de caso em questão tem-se somente a função “cortar”.

O cenário contava com dois objetos: um quadrado, de material reciclado, e uma bola de borracha (Figura 14). Primeiro ambos eram apresentados ao voluntário e então era informado a ele que ambos os objetos seriam lançados de um lado para o outro em diferentes sentidos e trajetórias e era pedido ao voluntário para simular o corte dos objetos quando da aproximação deles (Figura 15). Como no contexto da aplicação o local da tela onde a função foi ativada importa, era necessário um gesto dinâmico, portanto esse cenário criado para a segunda etapa estimulou esse tipo de gesto. Todos os participantes autorizaram a filmagem dos experimentos

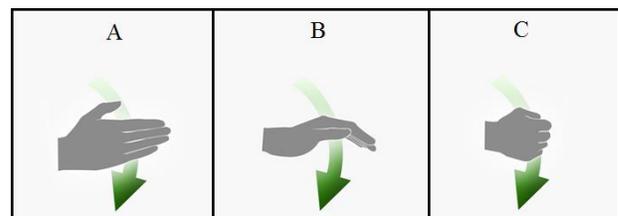
Na terceira etapa, foram analisadas as gravações obtidas anteriormente e constatou-se que dentre os dez gestos obtidos havia três variações, gesto A (Figura 16A), B (Figura 16B) e C (Figura 16C), onde o A foi executado por cinco voluntários, o B por quatro e o C por um. Por serem poucas opções de gestos, decidiu-se selecionar todas as três para a próxima etapa a fim de prover os voluntários com a maior diversidade possível.



**Figura 14 - Objetos utilizados no cenário da segunda etapa**



**Figura 15 - Voluntários atuando no cenário da segunda etapa**



**Figura 16 - Três variações de gestos**

Na quarta etapa, foi feita uma entrevista com os voluntários da segunda etapa, na qual eram expostas as sugestões dos três gestos selecionados na anterior, sendo que ainda seria possível escolher

outro totalmente diferente. Nenhum gesto novo foi sugerido. Os voluntários escolheram o gesto A como mais apropriado para a função “cortar” no cenário proposto. No fim desta etapa, o gesto A foi escolhido por sete voluntários, o B por dois e o C por um e não houve novas sugestões. Para a quinta etapa foi selecionado dois vocabulários de gestos, um com o A outro com o B.

Na quinta etapa, o processo deve avaliar os vocabulários unitários de gestos através de três testes: teste de atribuição de semântica, de memória e de stress. Porém, devido aos vocabulários possuírem apenas um gesto cada, somente faz sentido fazer o teste de stress, mas para ilustrar será descrito como seria o passo-a-passo dos outros testes.

No teste de atribuição semântica, o voluntário receberia um *template*, conforme a Figura 16, com os gestos, no caso o gesto A, e a lista de funções “cortar”. Caberia então ao voluntário fazer a correspondência correta entre gestos e funções. Em seguida, seria repetido o processo com o próximo vocabulário de gestos, no caso o B. A pontuação seria a soma dos palpites errados - zero, pois não teria como errar - dividido pelo número de gestos, um. Sendo assim, a pontuação seria sempre zero. Esse teste não foi executado por voluntários, essa descrição é meramente ilustrativa.

No teste de memória, seria mostrado aos voluntários um dos vocabulários de gestos, por exemplo o A, depois um slide com as funções sendo 2 segundos por função, no caso a função “cortar”. O participante deve então fazer o gesto correspondente a cada função até obter todas as corretas, caso erre, o procedimento do slide é reiniciado. Já que não haveria como errar, não haveria nenhuma reinicialização. O vocabulário de gestos seria revisado entre cada reinicialização do slide, aqui a pontuação corresponderia ao número de reinicializações, no presente caso a pontuação seria zero novamente. Esse teste também não foi executado por voluntários, essa descrição é meramente ilustrativa.

O teste de stress foi executado com voluntários, nele cada voluntário repetiu cada gesto trinta vezes e então foram questionados quão estressantes era cada um, fornecendo uma classificação geral para cada gesto usado durante um longo tempo; “sem problema”, “ligeiramente cansativo”, “irritante”, “doloroso” ou “impossível”. O gesto A obteve oito “sem problema” e dois “ligeiramente cansativo”, o B obteve sete “sem problema” e três “ligeiramente cansativo”. No caso em estudo, o gesto A obteve uma classificação ligeiramente melhor que o B, suas pontuações empatariam.

#### 4. CONSIDERAÇÕES FINAIS

O piloto apresentado serve apenas pra exemplificar a aplicação, do estado atual, do processo de desenvolvimento de gestos. A proposta deste trabalho é desenvolver uma aplicação piloto cujos gestos sejam determinados a partir do processo aqui descrito. Para o desenvolvimento das aplicações piloto é necessário definir um dispositivo para prover a interação natural, escolheu-se o Leap Motion. Na sua versão final, as aplicações serão então avaliadas e testadas com usuários reais no intuito de validar indiretamente seu vocabulário de gestos e também o processo utilizado no seu desenvolvimento.

#### 5. REFERÊNCIAS

[1] FUTURELAB. Interação Natural. Disponível em: <http://www.futurelab.com.br/site/interacao-natural/>, 2014.

[2] SBC. Sociedade Brasileira de Computação. Disponível em: [http://www.sbc.org.br/index.php?option=com\\_content&view](http://www.sbc.org.br/index.php?option=com_content&view)

=category&layout=blog&id=45&Itemid=66, 2014.

[3] Valli, A. *The Design of Natural Interaction*. 2006.

[4] McCord, Charlotte. *Gestures, Western Folklore*. 1948.

[5] Almeida, F. B. *Sistema Interativo Baseado em Gestos para Utilização de Comandos no Computador*. UFPE, 2013.

[6] Nielsen, M., Störring, M., Moeslund, T. B., and Granum, E. *A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for Man-Machine Interaction*. 1601-3646, Aalborg University, Aalborg, Dinamarca, 2003.

[7] Nielsen, M., Moeslund, T. B., Störring, M., and Granum, E. HCI Beyond the GUI: Design for Haptic, Speech, Olfactory and Other Nontraditional Interfaces. In Kortum, Phillip, ed., *HCI Beyond the GUI: Design for Haptic, Speech, Olfactory and Other Nontraditional Interfaces*. Morgan Kaufmann, 2008.

[8] Nielsen, M., Störring, M., Moeslund, T. B., and Granum, E. *A procedure for developing intuitive and ergonomic gesture interfaces for HCI*. Aalborg University, Aalborg, Dinamarca, 2004.

[9] Henze, N., Löcken, A., Boll, S., Hesselmann, T., and Pielot, M. *Free-Hand Gestures for Music Playback: Deriving Gestures with a User-Centred Process*. University of Oldenburg, OFFIS - Institute for Information Technology, Oldenburg, Alemanha, 2010.

[10] Epps, J., Lichman, S., and Wu, M. *A Study of Hand Shape Use in Tabletop Gesture Interaction*. National ICT Australia, Alexandria, 2006.

[11] Choi, E., Kwon, S., Lee, D., Lee, H., and Chung, M. K. *Towards successful user interaction with systems: Focusing on user-derived gestures for smart home systems*. Pohang University of Science & Technology, 2014.

[12] Bodiroza, S., Stern, H.I., and Edan, Y. Dynamic gesture vocabulary design for intuitive human-robot dialog. In *7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Boston, MA 2012), 111 - 112.

[13] Ma, Y., Paterson, H. M., and Pollick, F. E. *A motion capture library for the study of identity, gender, and emotion perception from biological motion*. University of Glasgow, Glasgow, Scotland, 2006.

[14] Hwang, B.-W., Kim, S., and Lee, S.-W. *A full-body gesture database for automatic gesture recognition*. 0-7695-2503-2, Korea University, Seoul, Korea, 2006.

[15] Bacim, F., Nabiyouni, M., and Bowman, D. A. *Slice-n-Swipe: A Free-Hand Gesture User Interface for 3D Point Cloud Annotation*. Virginia Tech, Minnesota, 2014.

[16] Jeong, S., Jin, J., Song, T., Kwon, K., and Jeon, J. W. *Single-Camera Dedicated Television Control System using Gesture Drawing*. Sungkyunkwan University, Suwon, South Korea, 2012.

