

Estudos com Observadores para Avaliação de Qualidade: Influência do Contexto e do Tipo de Dados Avaliados

Samuel Silva
IEETA
Universidade de Aveiro
sss@ieeta.pt

Carlos Ferreira
DEGEI – Universidade de Aveiro
CIO – Universidade de Lisboa
cferreira@egi.ua.pt

Joaquim Madeira Beatriz Sousa Santos
DETI / IEETA
Universidade de Aveiro
{jmadeira,bss}@det.ua.pt

Resumo

Os estudos com observadores são uma possível forma para a validação e teste quer de interfaces de utilizador, quer de sistemas de visualização. Contudo, apesar das inúmeras directrizes propostas na literatura, devido a diversos factores inerentes a essas experiências, o seu desenho é um processo complexo. Com base nos resultados obtidos em dois estudos com observadores, para avaliação da qualidade de modelos poligonais, ilustra-se o efeito de dois factores particulares (contexto e tarefas não-isomórficas).

Palavras-Chave

Estudos com observadores, Malhas Poligonais, Qualidade

1. INTRODUÇÃO

Os estudos com observadores são, hoje em dia, uma prática frequente para a validação e teste de interfaces de utilizador e sistemas de visualização [Tea05]. Um relatório recente [Johnson 06] reafirma os processos de avaliação como fundamentais para determinar o sucesso e caracterizar a aplicabilidade de novas técnicas.

A literatura [Cochran 83, Dea04] tem proposto diversos conjuntos de regras que permitem o desenvolvimento de protocolos de avaliação para os mais variados fins, de forma a que estes possibilitem a recolha de informação relevante e, tanto quanto possível, independente de factores como a ordem das tarefas, aprendizagem e fadiga dos observadores. No entanto, existem factores que podem influenciar uma avaliação (introduzindo desvios nos resultados obtidos) e que muitas vezes não são considerados, tais como [North 06]:

- A escolha das tarefas e a forma como as questões são formuladas;
- O contexto proporcionado a uma tarefa pelas outras tarefas envolvidas na avaliação;
- No desempenho de várias tarefas, à partida equivalentes (isomórficas), as diferenças entre os dados avaliados em cada uma delas;
- Forçar os utilizadores a articular uma resposta pode tornar difícil o tratamento dos dados resultantes, mas

disponibilizar um conjunto de respostas possíveis pode levar o utilizador a decidir usando um processo de eliminação.

A variabilidade introduzida por estes factores torna o desenvolvimento de estudos com utilizadores um processo mais complexo e pode traduzir-se, por ex., na necessidade de aumentar o número de participantes de forma a conseguir maior significância estatística dos dados.

O presente artigo apresenta alguns resultados obtidos para a avaliação da qualidade percebida de modelos poligonais, através de duas experiências de avaliação com observadores, que permitem confirmar a importância de dois dos factores acima referidos.

Começa-se com uma breve descrição do protocolo utilizado para efectuar a avaliação da qualidade dos modelos poligonais de forma a permitir uma compreensão de todo o processo. Depois, alguns dos resultados obtidos são apresentados e discutidos. Finalmente, são apresentadas algumas conclusões.

2. ESTUDOS COM OBSERVADORES

Foi desenhada uma experiência com observadores de forma a avaliar a qualidade percebida de modelos poligonais. Esta experiência consistia em duas fases distintas:

- Fase 1 — Preferências — Os observadores eram con-

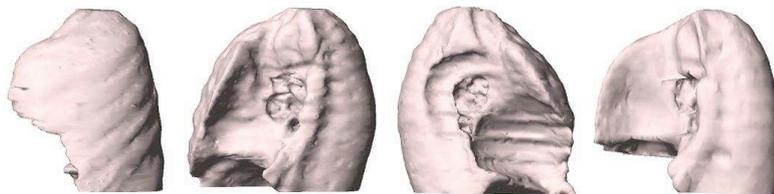


Figure 1. Modelos usados na primeira experiência (Conjunto A). O modelo pulmonar mais à direita foi também usado na segunda experiência.



Figure 2. Modelos usados na segunda experiência (Conjunto B). Do conjunto fazia também parte o modelo pulmonar apresentado à direita na figura 1.

frontados com o modelo original e várias versões do mesmo, resultantes de processos de simplificação usando diferentes algoritmos, para um mesmo nível de simplificação (20% e 50% do número original de faces) sendo-lhes pedido que ordenassem (1^o, 2^o ou 3^o) cada um dos modelos simplificados de acordo com a sua qualidade comparativamente ao original.

- Fase 2 — Classificações — Os observadores eram confrontados com um modelo original e uma sua versão simplificada sendo-lhes pedido que classificassem a qualidade da mesma comparativamente ao original usando uma escala de Likert de 5 níveis.

Durante estas duas fases foi também recolhida, para cada observador, informação referente ao número de interações com os modelos (rotações, translações, etc.) e ao tempo despendido para tomar as decisões.

A experiência foi executada para dois conjuntos distintos de modelos poligonais: um conjunto de modelos pulmonares (constituído pelos modelos apresentados na figura 1 — conjunto A) tendo participado 32 observadores; e, posteriormente, um conjunto de modelos de diferente natureza (constituído pelos modelos apresentados na figura 2 e também pelo modelo apresentado mais à direita da figura 1 — conjunto B) tendo participado 65 observadores.

Uma descrição mais detalhada do protocolo usado e dos resultados globais obtidos pode ser encontrada em [Silva 05, Silva 06].

3. RESULTADOS

O facto de os conjuntos de modelos usados em cada uma das experiências conterem um modelo comum permitiu re-

alizar a avaliação da sua qualidade enquadrando-o em diferentes contextos: no primeiro caso somente com modelos pulmonares (conjunto A), no segundo com modelos de diferente natureza (conjunto B). Os resultados apresentados foram obtidos usando o *software* STATISTICA [Sta06].

Começando pelos dados recolhidos durante a primeira fase das duas experiências, foi feita uma comparação dos tempos de decisão e do número de interações registados e referentes ao modelo pulmonar (e suas versões simplificadas). Para os tempos de decisão não se notou nenhuma diferença significativa. No entanto, para o número de interações, como se pode verificar pela análise dos diagramas de extremos e quartis [Hea83] apresentados na figura 3, e para ambos os níveis de simplificação (20% e 50%), o número de interações registadas é aproximadamente o dobro quando o modelo pulmonar se encontra inserido no conjunto B, do que na experiência em que apenas foram usados modelos pulmonares (conjunto A). A significância estatística deste resultado foi confirmada utilizando um teste não paramétrico de Mann-Whitney [Gibbons 97] ($p = 0.0004 \ll 0.05$ e $p = 0.002 \ll 0.05$ para os níveis de simplificação de 20% e 50% respectivamente) que rejeitou a hipótese nula (igualdade das medianas). De notar que não se registou nenhuma diferença significativa, entre as duas experiências, nas preferências atribuídas pelos utilizadores.

De seguida foram analisados os dados obtidos para o modelo pulmonar na segunda fase das experiências. O tempo de decisão foi menor para as situações em que o modelo pulmonar estava inserido no conjunto B (figura 4), tendo sido rejeitada a hipótese de igualdade de medianas através

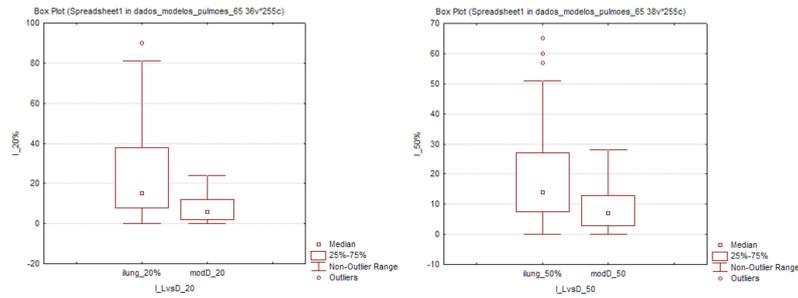


Figure 3. Número de interações registadas para o modelo comum, na fase das preferências, e para os níveis de simplificação de 20% (à esquerda) e 50% (à direita): em ambos os casos o diagrama da direita corresponde à experiência usando apenas modelos pulmonares (conjunto A).

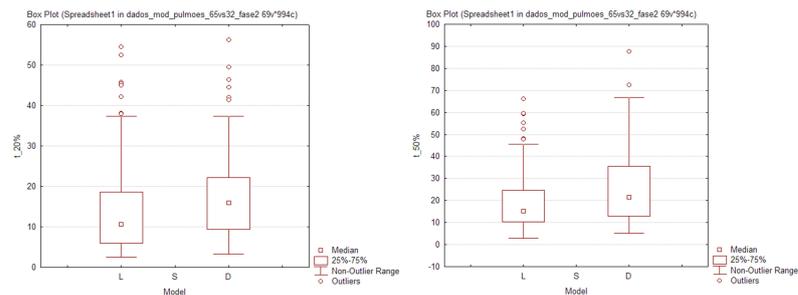


Figure 4. Tempos de decisão registados para o modelo comum, na fase das classificações, e para os níveis de simplificação de 20% (à esquerda) e 50% (à direita): em ambos os casos o diagrama da direita corresponde à experiência usando apenas modelos pulmonares (conjunto A).

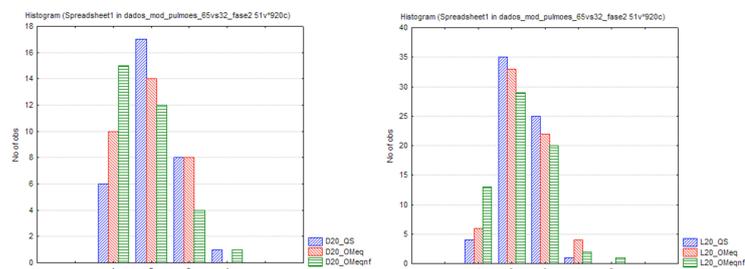


Figure 5. Histogramas das classificações obtidas pelas três versões do modelo pulmonar comum às duas experiências, para um nível de simplificação de 20%: o histograma da esquerda corresponde à experiência usando apenas modelos pulmonares (conjunto A).

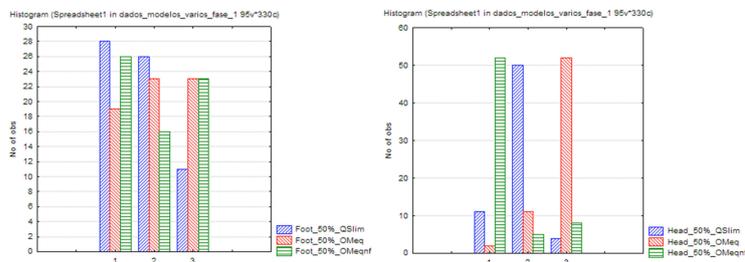


Figure 6. Histogramas das preferências atribuídas aos três algoritmos de simplificação avaliados para o modelo do pé (à esquerda) e para o modelo da cabeça (à direita), para um nível de simplificação de 50%.

de um teste de Mann-Whitney: $p = 0.000169 \ll 0.05$ e $p = 0.000281 \ll 0.05$ para os níveis de simplificação de 20% e 50%. Para o número de interações verificou-se uma igualdade para o nível de simplificação de 20%; para o nível de simplificação de 50% os observadores interagiram menos com o modelo pulmonar quando este se encontrava inserido no conjunto B.

Finalmente, comparando os histogramas das classificações obtidas pelas três versões simplificadas do modelo pulmonar comum em ambas as experiências (figura 5), verifica-se que os modelos simplificados obtiveram globalmente melhores resultados aquando da segunda experiência: o número de classificações mais baixas é proporcionalmente menor.

A análise dos dados referentes à segunda experiência [Silva 06] revelou um comportamento diferente do modelo da cabeça em relação aos outros modelos. Enquanto que para os outros modelos o tempo de decisão aumenta com o nível de simplificação, com a cabeça passa-se o contrário, sendo os observadores mais rápidos para o modelo menos simplificado. Para além disso, os histogramas das preferências do modelo da cabeça para os dois níveis de simplificação têm uma particularidade não observada nos dos outros modelos e que pode ser percebida, por exemplo, pela análise dos histogramas apresentados na figura 6: note-se que o histograma da cabeça (à direita) apresenta uma mais clara distinção entre os métodos de simplificação do que o referente ao modelo do pé (similar ao dos outros modelos).

4. CONCLUSÕES

Os resultados apresentados demonstram a importância de factores que nem sempre são considerados no desenho e análise dos resultados obtidos numa experiência com observadores. A análise dos dados resultantes da avaliação da qualidade do modelo poligonal do pulmão, obtidos nas duas experiências, permitiu verificar que, de facto, o diferente enquadramento deste nos dois conjuntos de modelos influenciou os resultados obtidos, nomeadamente ao nível dos tempos de decisão, do número de interações e das classificações obtidas. O facto de os observadores interagirem mais com o modelo pulmonar quando este estava inserido no conjunto A parece indicar uma análise mais cuidada por parte dos observadores (mais exigente) o que explicaria os valores mais baixos obtidos nas classificações. A presença de modelos de um único género parece ajudar a que os observadores foquem a sua atenção em detalhes dos mesmos (analisando-os com mais cuidado) que de outra forma passariam despercebidos.

Os resultados mais interessantes surgiram para o modelo da cabeça (segunda experiência) e demonstram um comportamento diferente deste modelo comparativamente aos outros. O tempo de decisão mais pequeno e a diminuição do mesmo para os modelos menos simplificados (contrariando os resultados globais obtidos de um tempo de decisão menor para modelos mais simplificados) podem ser motivados pela elevada capacidade do cérebro humano de processar informação de faces [Gregory 98] e pelo facto

de a análise de faces ser uma tarefa usual no dia a dia, o que treina o observador sobre as características principais a analisar. Para além disso, o modelo da cabeça permitiu discriminar, de forma mais clara, entre os métodos de simplificação, o que levanta a questão sobre quais as suas características que motivam este comportamento, e também se o modelo exibe a mesma capacidade de discriminação para outros métodos de processamento.

Apesar de obtidos num contexto particular e carecendo de uma análise mais cuidada, os resultados apresentados demonstram que é importante considerar os detalhes relacionados com o contexto em que as tarefas são desempenhadas, já que estes podem influenciar os resultados obtidos. Demonstram também que é necessário algum cuidado no desenho de tarefas que se pretendem isomórficas (diferentes instâncias do mesmo tipo de tarefa) sob pena de, apesar de formalmente equivalentes, apresentarem resultados influenciados pela natureza dos dados.

5. AGRADECIMENTO

O primeiro autor agradece à Unidade de Investigação 127/94 IEETA, da Universidade de Aveiro, a bolsa que vem financiando o seu trabalho.

BIBLIOGRAFIA

- [Cochran 83] W. G. Cochran. *Planning and Analysis of Observational Studies*. John Willey, 1983.
- [Dea04] A. Dix et al. *Human Computer Interaction*. Prentice Hall, 3ª edição, 2004.
- [Gibbons 97] J. D. Gibbons. *Nonparametric Methods for Quantitative Analysis*. American Sci. Press, 3ª edição, 1997.
- [Gregory 98] V. Gregory. *Eye and Brain, the Psychology of Seeing*. Oxford Univ. Press, 5ª edição, 1998.
- [Hea83] D. Hoaglin et al. *Understanding Robust and Exploratory Data Analysis*. John Wiley & Sons, 1983.
- [Johnson 06] C. Johnson, et al. *NIH / NSF Vis. Research Challenges Rep.* IEEE Press, 2006.
- [North 06] C. North. Toward measuring visualization insight. *IEEE Comp. Graph. & App.*, 26(3):pp. 6–9, 2006.
- [Silva 05] S. Silva, et al. Comparing three methods for simplifying mesh models of the lungs: an observer test to assess perceived quality. Em *Proc. SPIE 2005, vol 5749*, páginas 484–495, 2005.
- [Silva 06] S. Silva, et al. Perceived quality of simplified polygonal meshes: Evaluation using observer studies. Em *Proc. Ibero-American Symp. in Comp. Graph. SIACG06*, 2006.
- [Sta06] Statistica 6.0. <http://www.statsoft.com>, 2006.
- [Tea05] M. Tory et al. Evaluating visualizations: Do expert reviews work? *IEEE Comp. Graph. & App.*, 25(5):pp. 8–11, 2005.