

Avaliação de Interfaces em Software Educacional: Comparando Experiências em dois Protótipos Sucessivos

Eliane Regina de Almeida Valiati

Luciano Vargas Flores

Evandro Manara Miletto

Marcelo Soares Pimenta

Universidade Federal do Rio Grande do Sul

Instituto de Informática

Av. Bento Gonçalves, 9500, Bloco IV, CEP 91501-970

Porto Alegre, RS, Brasil

(051) 3316 - 6859

evaliati,lvf,miletto,mpimenta@inf.ufrgs.br

Abstract

This paper presents and discusses the results of two evaluation experiments, done with successive versions of a music education software, involving: user test, heuristic evaluation with interface experts and with experts in the application's knowledge domain, as well as pedagogical aspects evaluation following a script specific for music education software evaluation. Such experiments yielded a situation that's uncommon in most projects - the possibility to compare the results of two evaluations - and allowed the identification of differences and similarities that confirm interesting aspects about the process of evaluating educational software.

Keywords: usability evaluation techniques, educational software interfaces, prototype cycles

Resumo

Este artigo apresenta e discute os resultados obtidos em dois experimentos de avaliação, realizados com versões sucessivas de um software para educação musical, envolvendo: teste com usuários, avaliação heurística com especialistas em interfaces e com especialistas no domínio da aplicação, assim como avaliação de aspectos pedagógicos segundo um roteiro específico para avaliação de programas para educação musical. Estes experimentos propiciaram uma situação incomum na maioria dos projetos - a possibilidade de comparar resultados de duas avaliações - e permitiram identificar diferenças e similaridades que confirmam aspectos interessantes sobre o processo de avaliação de software educacional.

Keywords: técnicas de avaliação de usabilidade, interfaces de software educacional, ciclos de prototipação

1 Introdução

A ponderação dos problemas de usabilidade de uma interface pode ocorrer com o emprego de diferenciados métodos de avaliação disponíveis na literatura. Porém, alguns métodos mostram-se mais adequados que outros se considerarmos as particularidades envolvidas no projeto de certos tipos de interfaces. Este é o caso dos softwares educacionais, onde o processo de avaliação pressupõe a adoção de certos métodos, imprescindíveis para um completo diagnóstico da interface, como os ensaios de interação envolvendo alunos potenciais e a utilização de técnicas de inspeção especificamente projetadas para uso em softwares educacionais. Além disso, a avaliação de uma aplicação educacional requer a participação não apenas de especialistas em interfaces, mas de avaliadores especialistas no domínio de conhecimento envolvido em seu projeto [10]. Assim, o processo de avaliação de usabilidade em software educacional geralmente exige não apenas o emprego de diferentes técnicas de avaliação, mas principalmente a participação de diferentes tipos de avaliadores em um projeto multidisciplinar.

Sem dúvida, os métodos de avaliação tentam cobrir de maneira eficiente um certo conjunto de critérios e condições. Definir uma abordagem própria de avaliação pode ser uma solução prática e adequada quando métodos, aplicados isoladamente e em sua íntegra, não atendem às expectativas e necessidades de determinados tipos de projeto. Nesses casos, os métodos podem ser estendidos, especializados e utilizados de forma complementar.

Neste contexto, o presente artigo busca fornecer um panorama de alguns métodos utilizados em situações reais de avaliação experienciadas no Laboratório de Computação e Música da UFRGS (LC&M), na elaboração do software para educação musical STI - Sistema para Treinamento de Intervalos [3] [1]. O STI é uma aplicação hipermídia para auxiliar o ensino da teoria dos intervalos musicais melódicos (tópico da teoria musical) a estudantes de música. Mais do que conteúdo teórico, esse sistema apresenta exercícios práticos, como ditado intervalar e assimilação de intervalos, que fazem parte do ensino regular de música, além de fornecer o acompanhamento do desempenho do aluno. Seu desenvolvimento segue uma proposta centrada no usuário, adaptada a projetos de software educacional [12], onde num “ciclo de prototipação” sucessivos protótipos são criados e avaliados, evoluindo até o produto final.

Este artigo apresenta, compara e discute os resultados de dois experimentos de avaliação da usabilidade de protótipos sucessivos do STI. Estes experimentos propiciaram uma situação incomum na maioria dos projetos - a possibilidade de comparar resultados de duas avaliações - e permitiram identificar diferenças e similaridades que confirmam aspectos interessantes sobre o processo de avaliação e redesign de software educacional.

O artigo está estruturado como segue. A seção 2 apresenta o primeiro experimento realizado com o STI, descrevendo o protótipo, os testes aplicados e os resultados obtidos; a seção 3 expõe, simetricamente, o protótipo, os testes aplicados e os resultados encontrados no segundo experimento realizado com a nova versão do STI, reprojeta após a primeira avaliação; a seção 4 compara e discute os resultados obtidos em ambos experimentos. A seção 5 contém as conclusões e considerações finais.

2 Experimento de Avaliação 1: STI versão Macintosh

O primeiro experimento com o software STI fez parte de uma investigação conduzida por Winckler et al. [11], com a intenção de verificar se os consultores do domínio de conhecimento específico da aplicação (no caso, música), treinados para realizarem o teste de avaliação heurística [9], poderiam aumentar a quantidade e melhorar a qualidade de suas sugestões nos projetos. Também buscou-se verificar que tipo de problemas de usabilidade seriam identificados por esses consultores (não especialistas em interfaces) e que relação tais problemas teriam com os problemas detectados pelos especialistas em interfaces utilizando o mesmo método. Maiores detalhes sobre essa investigação podem ser encontrados em [11].

Nesse trabalho os autores comentavam a dificuldade de se desenvolver software com usabilidade quando a participação de consultores do domínio de conhecimento tratado no software é necessária. Desenvolver projetos com uma equipe multidisciplinar é uma tarefa complexa, tanto pela dificuldade do trabalho cooperativo e de coordenação dos participantes, como por problemas de comunicação entre membros da equipe devido à diferença de domínios de conhecimento de cada um [7].

2.1. Protótipo

O protótipo utilizado nesse experimento foi a primeira versão do STI [3]. Esse protótipo foi desenvolvido em HyperCard, para execução no ambiente Apple Macintosh. A versão do HyperCard utilizada ainda não gerava interfaces coloridas.

2.2. Métodos de avaliação utilizados

Para a condução de sua investigação, Winckler et al. [11] propuseram três testes, cujos resultados foram comparados aos problemas encontrados originalmente pela equipe de projetistas e consultores sem o uso de técnicas de usabilidade:

- Teste com usuários, na forma de ensaios de interação [8];
- Avaliação heurística com especialistas em interfaces; e
- Avaliação heurística com consultores em música.

Nos ensaios de interação participaram voluntariamente cinco (5) alunos da Escola de Música Prelúdio (Porto Alegre, RS), com idades entre 13 e 15 anos, considerados representativos do público-alvo desta aplicação. O teste foi conduzido no ambiente real de estudo dos usuários, utilizando o método “Think Aloud” simplificado [8] em sessões individuais não filmadas. Cada sessão durou em média 25 minutos. Os usuários deveriam completar uma lista de 22 tarefas com o sistema e, no final do teste, comentar suas impressões sobre a interface.

A avaliação heurística com especialistas em interfaces contou com três (3) avaliadores, especialistas em Interação Humano-Computador (IHC), já familiarizados com a técnica. As três sessões individuais foram filmadas e duraram em média 37 minutos.

Na avaliação heurística com consultores em música também participaram três (3) avaliadores, que foram treinados previamente nessa técnica de avaliação. As três sessões individuais foram filmadas e duraram em média 65 minutos.

Após as sessões os problemas encontrados foram classificados pelos avaliadores de acordo com as heurísticas de usabilidade que violaram. Observa-se que as 10 heurísticas utilizadas, baseadas nas de [9], não foram suficientes para classificar problemas do domínio de conhecimento específico música. Heurísticas adicionais foram então criadas pelos consultores em música para satisfazer esta condição.

Os problemas foram também classificados por todos avaliadores em uma escala de severidade especial, diferente da proposta em [9] pois leva em consideração a influência sobre a realização das tarefas (Tabela 1).

Tabela 1. Escala de severidade de problemas de usabilidade

Severidade	Descrição
0	Problema cosmético, pouca importância para a execução da tarefa.
1	Afeta levemente a execução da tarefa.
2	Causa confusão ao usuário e atrapalha sensivelmente a execução da tarefa.
3	O usuário fica muito confuso ou completa a tarefa com muita dificuldade.
4	O usuário não consegue completar ou desiste da tarefa por causa do problema.

2.3. Resultados

Um resultado importante encontrado nesse primeiro experimento foi que a participação dos consultores nas avaliações possibilitou a identificação de problemas do domínio de conhecimento específico tratado no software, e não apenas de problemas de usabilidade.

Segundo Kay [5], para o usuário, a interface é o programa. Isto é, o usuário constata através da interface os problemas tanto de usabilidade quanto funcionais. Então existem problemas que são encontrados nas avaliações mas não são problemas de usabilidade ou de interface, mas de funcionalidade e conteúdo do sistema.

Foram considerados “problemas do domínio da música” aqueles que faziam referência explícita a conteúdos da teoria musical como, por exemplo, problemas relacionados ao tamanho e complexidade das melodias apresentadas no software. Todos os problemas encontrados passaram, então, a ser classificados como problemas de música ou de interface em ambos experimentos aqui descritos.

A Tabela 2 resume as quantidades de problemas encontrados por cada grupo de avaliadores e as intersecções entre eles. Os totais da coluna 4 referem-se ao número de problemas diferentes encontrados por cada grupo de

avaliadores, ou seja, foram descontadas as repetições dentro de cada grupo (problemas iguais identificados por avaliadores diferentes no mesmo grupo). A coluna 5 refere-se às quantidades de problemas encontrados exclusivamente por cada grupo de avaliadores, ou seja, descontadas as repetições entre grupos diferentes (as intersecções).

Tabela 2. Quantidades de problemas encontrados no primeiro experimento, por grupo de avaliadores e intersecções entre eles.

Grupo de Avaliadores	Problemas de interface	Problemas de música	Total	Exclusivamente encontrado por cada grupo	Em comum apenas com usuários	Em comum apenas com especialistas em IHC	Em comum apenas com especialistas em música
Usuários	36	0	36	17	-	11	6
Especialistas em IHC	54	0	54	34	11	-	7
Especialistas em música	21	30	51	36	6	7	-
Total	111	30	141	87			

O número de problemas em comum encontrados por todos os grupos de avaliadores foi 2, o que significa que houve pouca intersecção nos resultados de avaliação.

O número total de problemas diferentes encontrados nas avaliações foi 113, o que corresponde aos 87 problemas encontrados exclusivamente por cada grupo somados aos 26 (= 11 + 6 + 7 + 2) problemas em comum entre os grupos. Destes 113 problemas 30 são de música e 83 de interface.

Portanto, comparando os resultados obtidos nos três testes deste primeiro experimento (ver gráficos da Figura 1), os consultores foram responsáveis em encontrar 100% dos problemas do domínio música. Porém, com relação à interface os consultores foram capazes de detectar, exclusivamente, apenas 7% dos problemas, os especialistas em interfaces encontraram 42% e os usuários 20%. Os restantes 31% foram os encontrados por dois ou mais grupos de avaliadores (porcentagens aproximadas).

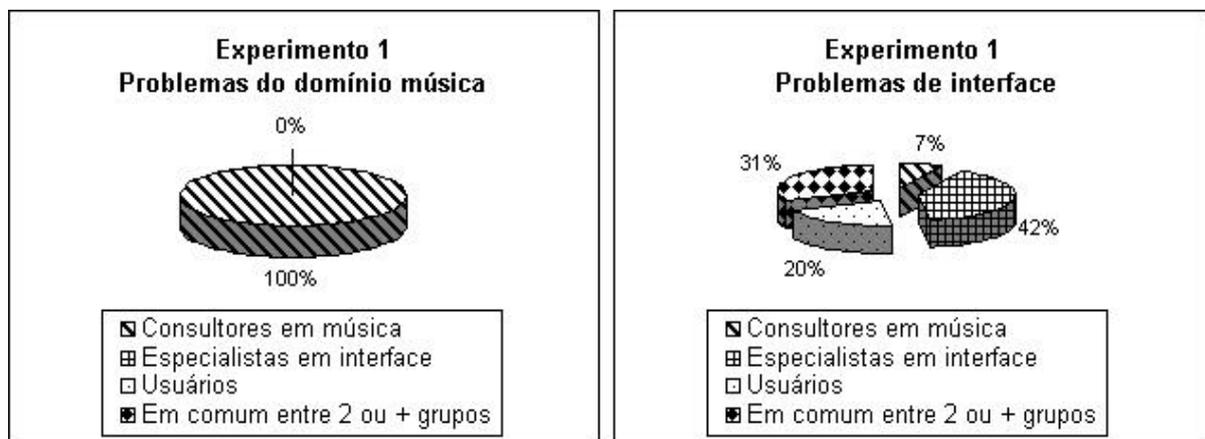


Figura 1. Distribuição dos problemas encontrados (em ambos os domínios) por diferentes grupos de avaliadores, no primeiro experimento.

3 Experimento de Avaliação 2: STI versão Windows

O segundo experimento visou a avaliação de um novo protótipo do STI, reprogramado para uma nova plataforma (Windows 95/98).

Alguns dos resultados do primeiro experimento foram levados em consideração nesta reengenharia do segundo protótipo do STI ("STI para Windows" [1]) e sua avaliação procurou reproduzir os mesmos testes do primeiro experimento. Um objetivo secundário deste experimento foi, portanto, comparar os resultados dos dois

experimentos para verificar se as conclusões seriam as mesmas. A opção por reproduzir os mesmos testes visava igualmente obter dados que fossem quantitativamente comparáveis, de modo a constatar se houve evolução de um protótipo para o seguinte devido ao emprego mais intensivo de técnicas de engenharia da usabilidade.

Foi realizada neste segundo experimento uma avaliação complementar do ponto de vista pedagógico, seguindo para isto um roteiro proposto por Krüger [6] específico para avaliação de software de educação musical. A avaliação foi feita por 3 educadores musicais, mas os resultados dessa avaliação pedagógica não serão comentados neste artigo pois representam uma avaliação muito mais qualitativa do que quantitativa, uma vez que são baseados na análise de comentários realizados pelos avaliadores e registrados no formulário do roteiro.

3.1. Protótipo

O protótipo utilizado nesse experimento foi a segunda versão do STI [1], desenvolvida em ToolBook II para execução na plataforma Windows 95/98 (mudanças em relação à primeira versão são comentadas adiante, na Seção 4.3).

3.2. Métodos de avaliação utilizados

De modo a reproduzir os procedimentos utilizados no primeiro experimento de avaliação, nesta segunda experiência também foram feitos três testes de avaliação da interface:

- Ensaios de interação. Aqui também foram observados 5 usuários pelo método “Think Aloud” simplificado, em sessões individuais não filmadas. As sessões duraram em média 34 minutos, durante os quais os estudantes completavam uma lista de 18 tarefas com o sistema e ao final comentavam suas impressões gerais;
- Avaliação heurística com 3 especialistas em interfaces; e
- Avaliação heurística com 3 consultores em música (professores de música).

Desta vez as avaliações heurísticas não puderam ser filmadas. As sessões individuais com os especialistas em interfaces duraram em média 45 minutos e as com consultores em música em média 82 minutos.

Os problemas identificados nas avaliações heurísticas também foram classificados de acordo com a heurística que violaram e com a mesma escala de severidade de problemas de usabilidade da Tabela 1.

3.3. Resultados

A Tabela 3 resume as quantidades de problemas encontrados por cada grupo de avaliadores e as intersecções entre eles.

Tabela 3. Quantidades de problemas encontrados no segundo experimento, por grupo de avaliadores e intersecções entre eles.

Grupo de Avaliadores	Problemas de interface	Problemas de música	Total	Exclusivamente encontrado por cada grupo	Em comum apenas com usuários	Em comum apenas com especialistas em IHC	Em comum apenas com especialistas em música
Usuários	22	4	26	12	-	6	5
Especialistas em IHC	39	0	39	26	6	-	4
Especialistas em música	15	18	33	21	5	4	-
Total	76	22	98	59			

O número de problemas em comum encontrados por todos os grupos de avaliadores foi 3, o que significa que, novamente, houve pouca intersecção nos resultados de avaliação.

O número total de problemas diferentes encontrados nas avaliações foi 77, o que corresponde aos 59 problemas encontrados exclusivamente por cada grupo somados aos 18 (= 6 + 5 + 4 + 3) problemas em comum entre os grupos. Destes 77 problemas 18 são de música e 59 de interface.

Analisando os resultados obtidos nos três testes deste segundo experimento (ver gráficos da Figura 2), comprova-se os dados encontrados no primeiro experimento, ou seja, os consultores novamente foram responsáveis em encontrar 100% dos problemas do domínio música, apesar da intersecção de 22% com problemas apontados pelos usuários (de qualquer modo, todos esses 4 problemas ainda foram identificados pelos consultores). Contudo, no que diz respeito a problemas de usabilidade na interface os consultores foram capazes de detectar, exclusivamente, apenas a parcela de 12%, os especialistas em interfaces encontraram 44% e os usuários 20%. Os 24% restantes correspondem aos problemas em comum, encontrados por dois ou mais grupos de avaliadores (porcentagens aproximadas).

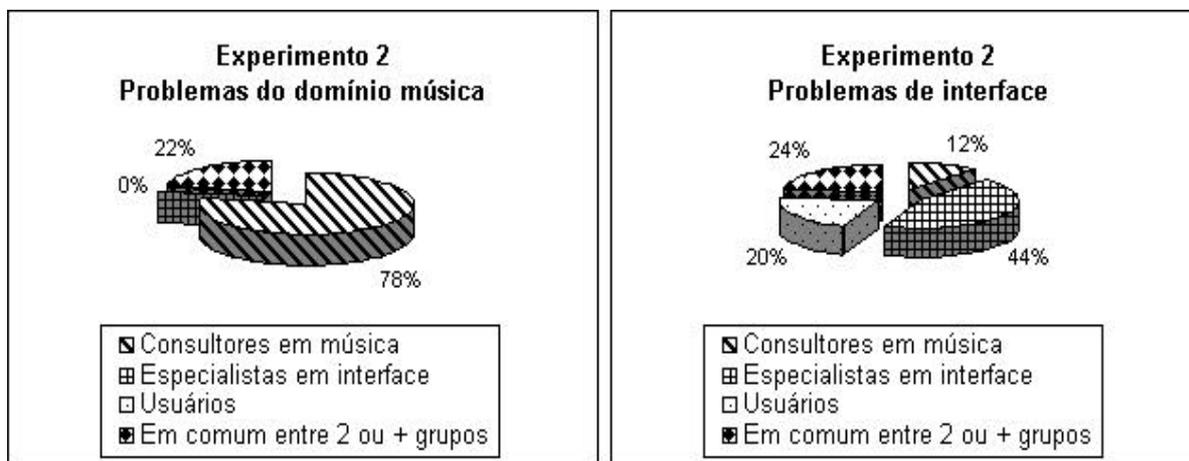


Figura 2. Distribuição dos problemas encontrados (em ambos os domínios) por diferentes grupos de avaliadores, no segundo experimento.

4 Discussão

Nesta seção são discutidos aspectos relacionados a equipes de avaliação, métodos utilizados e mudanças observadas nos dois protótipos implementados. Também são analisados, comparados e discutidos os resultados gerais obtidos em ambos experimentos de avaliação, realizados nos dois protótipos sucessivos do software STI.

4.1. Equipe de avaliação

Nos dois experimentos foram envolvidos o mesmo número de avaliadores especialistas em interface, consultores em música e usuários. Porém, cabe salientar que as pessoas participantes de cada grupo de avaliadores não foram as mesmas no primeiro e no segundo experimento.

Contudo, apesar de terem atuado pessoas diferentes, o desempenho geral apresentado por cada grupo de avaliadores (em cada domínio) permaneceu praticamente constante em ambos experimentos (conforme gráficos das Figuras 1 e 2). Nota-se apenas que houve um relativo acréscimo, do primeiro experimento para o segundo, na média de tempo gasto pelos avaliadores nos três testes realizados (usuários média de 25 min para 34 min, especialistas em interface média de 37 min para 45 min e consultores em música média de 65 min para 82 min, respectivamente).

4.2. Métodos utilizados

Durante a avaliação dos dois protótipos foram empregados, basicamente, os mesmos tipos de testes, ou seja, ensaios de interação, avaliação heurística por especialistas em interface e por consultores em música.

As únicas diferenças observadas, no segundo experimento, foram que nenhuma das sessões de avaliação foi filmada e que foi realizada uma avaliação pedagógica adicional (não considerada nesta discussão).

4.3. Mudanças observadas nos protótipos

A segunda versão resultou da reengenharia do STI original (para Apple Macintosh) visando sua adaptação à plataforma Windows 95/98 [1]. Foram mantidas as mesmas funcionalidades do STI para Macintosh (conteúdo teórico, exercícios práticos e acompanhamento do desempenho do aluno), apesar de algumas modificações na interface gráfica, devidas às diferenças que esta apresenta na nova plataforma.

No projeto do STI para Windows procurou-se corrigir alguns dos problemas identificados no primeiro experimento de avaliação, principalmente os relativos a controle e orientação do usuário, sobrecarga de informação e consistência da interface, mas isso foi feito de forma pouco sistemática (o que será discutido em artigo futuro - ver Seção 5).

4.4. Resultados

O número total de problemas diferentes encontrados no segundo experimento (77 problemas) foi menor que o encontrado no primeiro experimento (113 problemas). Com isso concluímos que a sistemática do grupo de pesquisa para o desenvolvimento de sistemas de educação musical de fato evoluiu, devido tanto à experiência prática já adquirida pelo grupo (ver Flores et al. [1]) como também à adoção de técnicas de engenharia de software e de usabilidade no projeto do segundo protótipo do STI.

Entretanto, a quantidade de problemas identificados nessa segunda avaliação ainda foi bastante expressiva. Devemos lembrar aqui que os resultados da primeira avaliação não foram eficientemente aproveitados no projeto do segundo protótipo, e que esse projeto incluiu a mudança de plataforma de execução. Esses dois fatores podem ter contribuído para a introdução de novos problemas de usabilidade no segundo protótipo do STI.

Mas este valor expressivo também pode significar que a avaliação foi realizada de forma bastante eficaz, o que confirma a afirmação em [11] de que a utilização de mais de uma técnica de avaliação mostra-se mais eficaz na prática do que o uso de uma só técnica. Para isto, basta comparar a quantidade total de problemas encontrados com a quantidade de problemas encontrados em cada um dos três testes. Por exemplo, no segundo experimento, se tivesse sido feita apenas a avaliação heurística com especialistas em interfaces, teríamos identificado 39 problemas, ao invés de 77.

Há outros resultados do segundo experimento que igualmente confirmam conclusões do primeiro experimento:

- *A importância de testes com usuários e com especialistas no domínio da aplicação*, neste caso a música, confirma-se pelas quantidades expressivas de problemas encontrados exclusivamente por esses dois grupos (no primeiro experimento, num total de 113 problemas, 17 foram encontrados pelos usuários e 36 pelos consultores, e no segundo experimento, do total de 77 problemas, foram detectados 12 por usuários e 21 por consultores). Resultados semelhantes também são encontrados no trabalho de [10], onde uma significativa quantidade de problemas foi identificada apenas por especialistas no domínio da aplicação.
- Com relação a problemas referentes à usabilidade da interface, os especialistas em interfaces foram capazes de encontrar uma quantidade maior de problemas neste domínio, por exemplo: no segundo experimento dos 26 problemas identificados pelos especialistas em interfaces todos eram referentes à usabilidade da interface, e dos 21 problemas encontrados pelos especialistas em música apenas 7 eram relativos a este domínio. Isso confirma que *“os consultores, mesmo quando treinados para o teste de avaliação heurística, não são tão eficientes quanto os especialistas em interface para encontrar problemas de usabilidade com a interface”* [11]. Ou seja, *especialistas no domínio da aplicação, treinados para a avaliação heurística, podem identificar problemas de usabilidade, mas não substituem os especialistas em interfaces*. Esta afirmação também é encontrada no trabalho de [10], pois seus experimentos comprovam que a participação dos especialistas em interfaces é imprescindível na detecção dos problemas referentes a usabilidade.
- Por outro lado, em ambos experimentos, problemas do domínio de conhecimento música foram identificados somente na avaliação heurística com os especialistas em música e no teste com usuários. Conforme gráficos das Figuras 1 e 2, constata-se novamente que *os especialistas em interfaces foram incapazes de identificar problemas do domínio de conhecimento específico* [11].

No primeiro experimento o teste com usuários não revelou nenhum problema do domínio da música. Na época isso foi explicado de duas formas: “primeiro porque há uma exigência de conhecimento para identificar tais tipos de problemas, que os alunos que participaram nos testes ainda não alcançaram; e, segundo, porque os alunos poderiam ter pensado que o software foi desenvolvido por professores de música e, portanto, não apresentava problemas com o conteúdo, e sentiram-se inseguros em comentar algum problema dessa natureza” [11]. Tais autores sugerem ainda que “com alunos mais avançados poderiam ser identificados problemas do domínio da teoria musical”. Essa sugestão se confirmou na segunda avaliação: os estudantes de música já conheciam um pouco mais sobre o conteúdo tratado no protótipo, sendo capazes de apontar 4 problemas do domínio da música. Observamos, no entanto, que a quantidade total de problemas de música encontrados nessa avaliação continua sendo 18, porque os 4 problemas encontrados no teste com usuários também foram

apontados pelos especialistas no domínio específico. Ou seja, o teste com usuários não identificou nenhum problema de música de forma exclusiva, que não fosse também identificado pelos especialistas em música. Isto aponta para a conclusão de que *usuários, assim como especialistas em interfaces, igualmente não são tão eficazes quanto os especialistas no domínio da aplicação para identificar problemas desse domínio na interface*, embora sejam sempre eficazes para identificar problemas de usabilidade.

5 Conclusão

A possibilidade de comparar os resultados das duas avaliações de usabilidade descritas neste artigo, por terem sido realizadas de forma semelhante, possibilitou confirmar a eficácia de alguns procedimentos recomendados por Valiati et al. [10] e Winckler et.al. [11] para uma melhor avaliação de interfaces em software educacional, entre eles:

- a) Usar mais de um método de avaliação de usabilidade para uma avaliação mais completa dos protótipos. Como nos testes houve pouca sobreposição de problemas encontrados, observa-se que cada método cobre problemas específicos com a interface e, assim, utilizados em conjunto obtém-se uma avaliação mais eficaz;
- b) No caso de se usar a técnica de avaliação heurística de usabilidade, certos problemas em sistemas educativos só podem ser identificados estendendo essa técnica para avaliar aspectos do domínio de conhecimento específico da aplicação. Para isso, contar com a participação de especialistas nesse domínio;
- c) Avaliar também outros aspectos, como o pedagógico (interesse despertado nos alunos, adequação do conteúdo e abordagem pedagógica utilizada, etc.). Apesar de não ter sido útil às comparações quantitativas deste artigo, observamos que o resultado geral da avaliação pedagógica no segundo experimento foi bastante satisfatório. Essa avaliação qualitativa, segundo um roteiro específico, mostrou-se de fato bastante válida, já que as críticas foram sempre acompanhadas de sugestões, de ótima qualidade, para a solução dos problemas apontados.

Na continuidade deste trabalho pretendemos efetuar mais iterações no ciclo de prototipação adotado para o projeto do STI [12] e prosseguir comparando os dados resultantes, discutindo e publicando-os. Também pretendemos investigar questões relativas ao aproveitamento adequado dos resultados da avaliação de cada protótipo no projeto do protótipo seguinte, já que tais questões igualmente surgiram no projeto aqui descrito.

Referências

- [1] Flores, L. V. *STI - Sistema para Treinamento de Intervalos para Plataforma Windows 95/98*. Projeto de Diplomação (Bacharelado em Ciência da Computação) - Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2000.
- [2] Flores, L. V.; Vicari, R. M.; Pimenta, M. S. *Some Heuristics for the Development of Music Education Software: First Steps Towards a Methodology*. In: *Brazilian Symposium on Computer Music*, 8.; Congresso da Sociedade Brasileira de Computação, 21., 2001, Fortaleza. **Anais...** Niterói: Instituto Doris Ferraz de Aragon. 1 CD. 2001.
- [3] Fritsch, E. F. *STI - Sistema para Treinamento de Intervalos*. In: *Simpósio Brasileiro de Computação e Música*, 3., 1996, Recife. **Anais...** Recife: Departamento de Música / UFPE. p.45-55. 1996.
- [4] Jeffries, R. et al. *User Interface Evaluation in the Real World: A Comparison of Four Techniques*. In: *ACM CHI'91 - Conference on Human Factors in Computing Systems*, 1991, New Orleans, Louisiana, EUA. **Proceedings...** New York: ACM Press. p.119-124. 1991.
- [5] Kay, A. *User Interface: A Personal View*. In: Laurel, B. (Ed.). *The Art of Human-Computer Interface Design*. Reading, EUA: Addison-Wesley. p.191-207. 1990.
- [6] Krüger, S. E. *Desenvolvimento, Testagem e Proposta de um Roteiro para Avaliação de Programas para Educação Musical*. Dissertação (Mestrado em Educação Musical) - Programa de Pós-Graduação em Música / UFRGS, Porto Alegre. 2000.
- [7] Lewis, C.; Rieman, J. *Task-Centered User Interface Design: A Practical Introduction*. Disponível em: <ftp.cs.colorado.edu/pub/distrib/clewis/HCI-Design-Book/>. Acesso em: maio/2002.
- [8] Nielsen, J. *Usability Engineering*. Mountain View, EUA: AP Professional. 1993.

- [9] Nielsen, J. *Heuristic Evaluation*. In: Nielsen, J.; Mack, R. L. (Ed.) Usability Inspection Methods. New York: John Wiley & Sons.1994.
- [10] Valiati, E. R. A.; Levacov, M.; Lima, J. V.; Pimenta, M. S. (2000) *Utilizando professores como avaliadores heurísticos de interfaces de software educacional*. In: SIIÉ'2000 - Simposio Internacional de Informática Educativa, 2., nov. 2000, Puertollano, Ciudad Real, Espanha. **Actas...** Ciudad Real: ADIE. 1 CD. 2000.
- [11] Winckler, M. A. A.; Nemetz, F.; Lima, J. V. *Estudo de Caso da Aplicação do Método de Avaliação Heurística em um Projeto Multidisciplinar*. In: IHC'98 - Workshop sobre Fatores Humanos em Sistemas Computacionais, 1., out. 1998, Maringá, Brasil. **Atas...** Rio de Janeiro: PUC-RJ. p.66-74. Disponível em: <protem.inf.ufrgs.br/lcm/download/pdfs/ihc98.pdf>. Acesso em: maio/2002.
- [12] Winckler, M. A. A.; Nemetz, F.; Lima, J. V. *Interação entre Aprendiz e Computador: Métodos para Desenvolvimento e Avaliação de Interfaces*. In: Tarouco, L. M. R. (Ed.) Tecnologia Digital na Educação. Porto Alegre: Pós-Graduação em Informática na Educação, UFRGS. p.7-33. 2000.