

**Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Engenharia de Produção**

Adriana Holtz Betiol

**AVALIAÇÃO DE USABILIDADE PARA OS
COMPUTADORES DE MÃO: UM ESTUDO COMPARATIVO
ENTRE TRÊS ABORDAGENS PARA ENSAIOS DE
INTERAÇÃO**

Tese de Doutorado

**Florianópolis
2004**

Adriana Holtz Betiol

**AVALIAÇÃO DE USABILIDADE PARA OS
COMPUTADORES DE MÃO: UM ESTUDO COMPARATIVO
ENTRE TRÊS ABORDAGENS PARA ENSAIOS DE
INTERAÇÃO**

**Tese apresentada ao Programa de
Pós-Graduação em Engenharia de
Produção da Universidade Federal de
Santa Catarina como requisito parcial
para obtenção do título de Doutor em
Engenharia de Produção**

Orientador: Prof. Walter de Abreu Cybis, Dr.

**Florianópolis
2004**

B563a Betiol, Adriana Holtz
Avaliação de usabilidade para os computadores de mão : um estudo comparativo entre três abordagens para ensaios de interação / Adriana Holtz Betiol; orientador, Walter de Abreu Cybis. – Florianópolis, 2004. 210 f. : il.

Tese (Doutorado) – Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia de Produção, 2004.

Inclui bibliografia.

1. Sistemas de comunicação móvel – Avaliação. 2. Telefone celular. 3. Interfaces (Computadores). 4. Interação homem-máquina. 5. Computação móvel. I. Cybis, Walter de Abreu. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia de Produção. III. Título.

CDU: 681.31

Catálogo na fonte por: Onélia Silva Guimarães CRB-14/071

Adriana Holtz Betiol

**AVALIAÇÃO DE USABILIDADE PARA OS
COMPUTADORES DE MÃO: UM ESTUDO COMPARATIVO
ENTRE TRÊS ABORDAGENS PARA ENSAIOS DE
INTERAÇÃO**

Esta tese foi julgada e aprovada para a obtenção do grau
de **Doutor em Engenharia de Produção** no **Programa de
Pós-Graduação em Engenharia de Produção** da
Universidade Federal de Santa Catarina

Florianópolis, 27 de fevereiro de 2004

Prof. Edson Pacheco Paladini, Dr.
Coordenador do Programa

BANCA EXAMINADORA

Profa. Leila Amaral Gontijo, Dra.
Universidade Federal de Santa Catarina

Prof. Walter de Abreu Cybis, Dr.
Universidade Federal de Santa Catarina
Orientador

Prof. Neri dos Santos, Dr.
Universidade Federal de Santa Catarina

Prof. Jean-Marc Robert, Dr.
École Polytechnique de Montréal

Prof. Bruno Farbos, Dr.
École Polytechnique de Montréal

Egberto Arouca Modesto de Medeiros, Dr.

A minha família,
com amor.

AGRADECIMENTOS

A Deus, por ter me proporcionado mais esta vitória.

Ao meu marido, por estar sempre ao meu lado, me incentivando e me apoiando a enfrentar novos desafios.

Aos meus filhos, minhas maiores conquistas.

Aos meus pais, por todo apoio e principalmente, por todo o tempo em que me ajudaram a cuidar das crianças para que eu pudesse me dedicar a este trabalho.

Aos meus irmãos, por eu saber que sempre poderei contar com vocês.

Ao meu orientador Prof. Walter de Abreu Cybis, por tudo que me ensinou sobre a Engenharia de Usabilidade. Seus ensinamentos e questionamentos foram decisivos em todas as etapas deste trabalho. Muito obrigada por sua orientação, amizade, disponibilidade e principalmente por ter acreditado em mim.

Aos membros da banca examinadora, pelo olhar crítico e competente, cujas contribuições e sugestões foram importantes para o aperfeiçoamento deste trabalho.

Aos professores do PPGEF da UFSC, por tudo que me ensinaram e me ajudaram a descobrir.

À Pontifícia Universidade Católica do Paraná, por ter permitido que eu realizasse este doutorado.

Ao pesquisador Didier Chincholle, por todo o material de pesquisa que me indicou, literatura de excelente conteúdo que muito me auxiliou neste trabalho.

A Edgar Alves, pela inestimável ajuda no projeto e na execução do tripé e do suporte da câmera sem fio.

Ao Prof. Flavio Bortolozzi da PUCPR, por sempre ter me incentivado em meus projetos.

Ao Prof. Mardson Freitas de Amorim da PUCPR, pelo imensurável apoio na preparação e montagem dos equipamentos do laboratório de testes.

Ao Prof. Marco Eleutério e sua equipe do CEAD da PUCPR, pela disponibilização de local e equipamentos para os testes.

Ao Prof. Ricardo Ehlers e Sandro Sartori da UFPR, pelo apoio durante a fase de análise estatística dos dados.

Meu agradecimento muito especial a todas as pessoas que com muita boa vontade e simpatia participaram dos ensaios de interação, sem vocês este trabalho não teria sido realizado.

“We have to break away from the tried-and-true methods of the computer and face up to design under new constraints, different environments, and restricted technology. The new world of interactive consumer-centered products poses interesting usage requirements. No longer can we assume a single user: the user maybe a group. No longer can we assume undivided attention: the user maybe driving an auto, walking down the street, or engaged in other activities. No longer can we assume a keyboard, a large screen, and a pointing device. Each product category poses novel challenges.”

Donald A. Norman

RESUMO

BETIOL, Adriana Holtz. **Avaliação de usabilidade para os computadores de mão: um estudo comparativo entre três abordagens para ensaios de interação.** 2004. 210f. Tese (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis.

A usabilidade dos computadores de mão e das aplicações para estes equipamentos é um fator determinante do sucesso deste segmento da computação móvel. As necessidades e características do usuário móvel, o contexto de uso dos computadores de mão, as características das aplicações e as limitações físicas destes equipamentos são fatores que irão influenciar a interação e devem ser considerados tanto no projeto das interfaces quanto na avaliação de usabilidade destes equipamentos. Esta pesquisa teve como objetivo investigar qual a influência do contexto de avaliação nos resultados da avaliação de usabilidade da interface com o usuário de uma aplicação de Internet móvel para telefone celular baseada na técnica de ensaios de interação. Os ensaios de interação foram realizados segundo três diferentes abordagens, utilizando: (i) emulador operado em laboratório, (ii) telefone celular operado em laboratório e (iii) telefone celular operado em campo. Os resultados do experimento mostraram que é possível obter bons resultados na avaliação de usabilidade de uma interface de um computador de mão utilizando um emulador no computador, desde que o contexto de avaliação seja definido da maneira mais próxima possível ao contexto de uso, assegurando principalmente a similaridade entre as interfaces. Também foi constatado que a posição do usuário e a maneira como ele segura um computador de mão são fatores que afetam a interação com a interface e devem ser considerados, assim como a realização de tarefas adicionais paralelas podem interferir nos resultados da usabilidade medida e nos índices de desempenho das avaliações. Novas técnicas que possibilitem ao usuário utilizar um computador de mão em situações reais, ou seja, em movimento, em vários ambientes diferentes, estando exposto às interferências externas, ao mesmo tempo que permitam registrar a interação são necessárias para auxiliar nas avaliações de usabilidade dos computadores de mão. Neste sentido, o sistema envolvendo uma minicâmera sem fio, desenvolvido especialmente para este trabalho, mostrou ser uma excelente solução para auxiliar na realização de ensaios de interação sempre que for necessário registrar as imagens e o áudio da interação do usuário com um computador de mão, independentemente do local da avaliação.

Palavras-chave: avaliação de usabilidade, computadores de mão, ensaios de interação, telefone celular, interação humano-computador

ABSTRACT

BETIOL, Adriana Holtz. **Usability evaluation of mobile devices: a comparative study of three approaches for usability testing**. 2004. 210f. Thesis (Doctorate in Production Engineering) – Post-Graduation Program in Production Engineering, UFSC, Florianópolis.

The usability of the mobile devices and their applications is a key factor for the success of mobile computing. The needs and characteristics of the mobile user, the usage context of the mobile devices, the characteristics of the applications and the physical limitations of these equipments are factors which will influence the interaction and should be considered in the design of the interfaces, as well as in the usability evaluation. The main objective of this research was to investigate the influence of the evaluation context in the results of usability tests conducted to evaluate the user interface of a mobile Internet application for mobile phones. The usability tests were conducted according to three different approaches: (i) using a computer-based mobile phone emulator inside the laboratory, (ii) using a mobile phone inside the laboratory and (iii) using a mobile phone in the field, in a more realistic environment. The results showed that it is possible to obtain satisfactory results in usability tests of a mobile user interface using a computer-based emulator, provided the evaluation context is defined as close as possible to the usage context, mainly assuring the similarity between the interfaces. It was also noticed that the position of the user and the way he holds the mobile phone are factors that affect the interaction with the interface and should be taken into consideration, as well as that the execution of concurrent tasks may interfere in the results of usability measured and problem identification. New techniques that allow the user to use a mobile device in real situations, that is, in movement, in many different places, being exposed to external interferences while registering the interaction are necessary to help in the usability evaluation of the mobile devices. In this sense, the wireless camera system, especially developed for this research, proved to be an excellent solution to record, in video and audio, the user's interaction with the mobile device, independently of where the evaluation takes place, either inside or outside the lab.

Keywords: usability evaluation, mobile devices, usability testing, mobile phones, human-computer interaction

SUMÁRIO

Lista de Figuras	xii
Lista de Quadros	xv
Lista de Tabelas	xvi
Lista de Siglas	xvii
INTRODUÇÃO	18
CAPÍTULO 1 – COMPUTADORES DE MÃO	29
1.1 Características dos computadores de mão	30
1.1.1 Telas pequenas.....	32
1.1.2 Entrada de dados.....	33
1.2 O usuário móvel e o contexto	37
CAPÍTULO 2 – USABILIDADE	40
2.1 Definições para usabilidade	42
2.1.1 Usabilidade segundo Shackel.....	42
2.1.2 Usabilidade segundo Nielsen.....	43
2.1.3 Usabilidade segundo Jordan.....	45
2.1.4 Usabilidade e as normas.....	46
2.1.4.1 ISO 9241-11.....	46
2.1.4.2 ISO/IEC 9126.....	47
2.1.4.3 ISO/IEC 14598.....	50
2.1.4.4 ISO 13407.....	51
2.1.4.5 ISO/TR 18529.....	53
2.1.5 Usabilidade segundo Hiltunen.....	54
2.2 Definição de critérios para medir a usabilidade	55
2.2.1 O projeto MUSiC.....	57
2.2.2 Etapas do Método de Medida da Performance.....	62
2.2.3 Medidas e métricas de usabilidade do MUSiC.....	64
2.2.3.1 Eficácia.....	66
2.2.3.2 Eficiência.....	67
2.2.3.2.1 Eficiência temporal.....	67
2.2.3.2.2 Eficiência relativa.....	67
2.2.3.3 Período produtivo.....	68

2.2.3.4 Medidas de satisfação.....	69
2.3 As técnicas de usabilidade no ciclo de desenvolvimento do produto.....	71
2.3.1 Ensaios de interação.....	73
2.3.1.1 Número de usuários em um ensaio de interação.....	77
2.3.1.2 Validade e confiabilidade dos ensaios de interação.....	79
2.4 Critérios para comparação entre técnicas de avaliação de usabilidade.....	80
CAPÍTULO 3 – AVALIAÇÃO DE USABILIDADE PARA OS COMPUTADORES DE MÃO.....	84
3.1 Avaliação de usabilidade com emuladores.....	84
3.2 Avaliação de usabilidade com o equipamento real.....	91
3.3 Avaliação de usabilidade fora do laboratório.....	100
3.4 Minicâmeras para avaliação de usabilidade dos computadores de mão.....	106
3.5 Prática nas empresas.....	108
3.6 Escolha das três abordagens para os ensaios de interação.....	110
CAPÍTULO 4 – DESCRIÇÃO DO EXPERIMENTO.....	112
4.1 Definição do produto a ser avaliado.....	112
4.2 Definição do contexto de uso.....	113
4.3 Definição dos diferentes contextos de avaliação.....	113
4.3.1 Contexto de avaliação 1: dentro do laboratório utilizando um emulador.....	114
4.3.2 Contexto de avaliação 2: dentro do laboratório utilizando o telefone celular e uma câmera de documentos.....	116
4.3.3 Contexto de avaliação 3: fora do laboratório utilizando o telefone celular e uma minicâmera sem fio.....	118
4.3.4 O emulador e o telefone celular.....	122
4.3.5 Definição da amostra de participantes.....	124
4.3.6 Definição do <i>script</i> das tarefas.....	126
4.3.7 Escolha das medidas e métricas de usabilidade.....	128
4.3.7.1 Usabilidade para os computadores de mão.....	128
4.3.7.2 Eficácia da tarefa.....	129

4.3.7.2.1 Tarefa completada.....	131
4.3.7.3 Eficiência do usuário.....	132
4.3.7.4 Eficiência relativa do usuário em relação ao usuário especialista.....	132
4.3.7.5 Período produtivo.....	132
4.3.7.6 Satisfação do usuário.....	134
4.3.7.7 Avaliação da carga de trabalho.....	135
4.4 Definição dos critérios para comparação dos desempenhos das avaliações.....	137
4.4.1 Atribuição de graus de severidade aos problemas de usabilidade.....	137
4.4.2 Definição das medidas e métricas de desempenho das avaliações.....	139
4.5 Condução dos ensaios de interação.....	139
4.5.1 Introdução do participante ao ambiente de teste.....	139
4.5.2 Treinamento do participante.....	140
4.5.3 Execução das tarefas.....	141
4.5.4 Questionários.....	141
4.5.5 Debriefing.....	142
4.6 Levantamento dos dados.....	142
4.7 Análise das fitas de vídeo.....	142
4.8 Análise dos problemas de usabilidade.....	144
4.9 Análise estatística.....	144
CAPÍTULO 5 – RESULTADOS.....	147
5.1 Usabilidade medida nos três contextos de avaliação.....	147
5.1.1 Eficácia da tarefa.....	147
5.1.2 Eficiência do usuário.....	151
5.1.3 Eficiência relativa do usuário em relação ao usuário especialista.....	153
5.1.4 Período produtivo.....	155
5.1.5 Satisfação do usuário.....	156
5.1.6 Avaliação da carga de trabalho.....	157
5.2 Medidas e métricas de desempenho das avaliações.....	159
5.2.1 Quantidade de problemas de usabilidade.....	159
5.2.2 Severidade dos problemas de usabilidade.....	162
5.2.3 Razão benefício/custo.....	164
5.3 Questionário ‘Perfil do Usuário’.....	165

5.4 Questionário 'Avaliação Geral sobre WAP'	169
CAPÍTULO 6 – CONCLUSÃO	175
6.1 Análise dos resultados	175
6.1.1 Usabilidade medida nos três contextos de avaliação.....	176
6.1.1.1 Eficácia, eficiência e satisfação.....	176
6.1.1.2 Período produtivo e eficiência relativa do usuário.....	178
6.1.1.3 Carga de trabalho.....	179
6.1.2 Desempenho das avaliações nos três contextos de avaliação.....	180
6.2 Discussão	181
6.3 Conclusão	185
6.4 Recomendações para trabalhos futuros	190
REFERÊNCIAS	192
APÊNDICE A – Roteiro com as informações fornecidas ao participante	201
APÊNDICE B – Questionário 'Perfil do Usuário'	202
APÊNDICE C – Instrumento para acompanhar a avaliação	206
APÊNDICE D – Questionário 'Avaliação Geral sobre WAP'	207
APÊNDICE E – Instrumento para acompanhar a análise da fita de vídeo	208
ANEXO A – Questionário de Satisfação SUS	209

LISTA DE FIGURAS

Figura 1: Número de usuários de telefonia móvel e Internet no mundo.....	19
Figura 2: Planta móvel brasileira (em milhões de usuários).....	19
Figura 3: Fatores que promovem o incremento da mobilidade.....	20
Figura 4: O computador pessoal.....	29
Figura 5: Exemplos de telas pequenas: PDA e telefone celular.....	32
Figura 6: Exemplo de tela pequena: relógio de pulso.....	33
Figura 7: Teclado Fastap.....	34
Figura 8: Minitclado QWERTY em PDA e telefone celular.....	35
Figura 9: Telefone celular com tela de toque.....	35
Figura 10: Teclado virtual por infravermelho.....	36
Figura 11: Teclados virtuais com sensores.....	36
Figura 12: Modelo de atributos de aceitabilidade de um sistema segundo Shackle.....	43
Figura 13: Modelo de atributos de aceitabilidade de um sistema segundo Nielsen.....	44
Figura 14: Modelo de qualidade para um produto de software ISO/IEC 9126-1.....	49
Figura 15: Processo de avaliação de produto de software ISO/IEC14598-1.....	51
Figura 16: Ciclo de projeto centrado no usuário ISO 13407.....	52
Figura 17: As normas e a usabilidade.....	54
Figura 18: Componentes da experiência do usuário móvel.....	54
Figura 19: Modelo de medida da usabilidade segundo Nielsen.....	56
Figura 20: Componentes da usabilidade.....	60
Figura 21: Etapas do Método de Medida da Performance.....	63
Figura 22: Relações entre as medidas e métricas de usabilidade no MUSiC.....	66
Figura 23: Exemplo de escala Likert.....	70
Figura 24: Exemplo de escala de diferencial semântico.....	70
Figura 25: Avaliação de usabilidade no desenvolvimento de um produto.....	72
Figura 26: Modelo de <i>layout</i> de um laboratório de usabilidade.....	75

Figura 27: Exemplo de emulador no computador.....	85
Figura 28: Usuário utilizando um emulador no computador.....	88
Figura 29: Câmeras de documentos.....	92
Figura 30: Ensaio de interação utilizando uma minicâmera sem fio.....	108
Figura 31: Países que responderam à pesquisa.....	109
Figura 32: Ferramentas utilizadas nos testes de usabilidade.....	109
Figura 33: Local dos testes de usabilidade.....	110
Figura 34: Componentes dos contextos de avaliação.....	114
Figura 35: Ambiente do contexto de avaliação do emulador.....	115
Figura 36: Imagem da interação no contexto de avaliação do emulador.....	115
Figura 37: Tripé para fixação do telefone celular.....	116
Figura 38: Ambiente do contexto de avaliação da câmera de documentos....	117
Figura 39: Imagem da interação no contexto de avaliação da câmera de documentos.....	117
Figura 40: Sistema da minicâmera sem fio.....	120
Figura 41: Sistema de recepção e gravação.....	121
Figura 42: Ambiente do contexto de avaliação da minicâmera sem fio.....	121
Figura 43: Imagem da interação no contexto de avaliação da minicâmera sem fio.....	122
Figura 44: Emulador e telefone celular.....	123
Figura 45: Diferenças entre as telas do emulador e do telefone celular.....	124
Figura 46: Árvore de decisão para categorizar as ações do usuário em relação ao tempo de execução da tarefa.....	133
Figura 47: Resultados para a eficácia da tarefa.....	147
Figura 48: Percentual dos usuários que completaram a tarefa.....	148
Figura 49: Telas da tarefa 1 no emulador e no telefone celular.....	149
Figura 50: Resultados para a eficiência do usuário.....	152
Figura 51: Resultados para a eficiência relativa do usuário em relação ao usuário especialista.....	153
Figura 52: Resultados para o período produtivo.....	155
Figura 53: Resultados para o questionário SUS.....	157
Figura 54: Resultados para o NASA TLX.....	158
Figura 55: Percentual de problemas encontrados por contexto de avaliação	160

Figura 56: Número total de problemas de usabilidade distribuídos por contexto de avaliação.....	161
Figura 57: Total dos graus de severidade dos problemas de usabilidade encontrados por contexto de avaliação.....	162
Figura 58: Freqüência dos problemas de usabilidade distribuídos por impacto.....	163
Figura 59: Razão benefício/custo por contexto de avaliação.....	165
Figura 60: Grau de instrução dos participantes.....	166
Figura 61: Ocupação dos participantes.....	166
Figura 62: Tempo como usuário da Internet.....	167
Figura 63: Freqüência de acesso à Internet.....	167
Figura 64: Média de tempo de acesso à Internet.....	168
Figura 65: Tempo como usuário de e-mail.....	168
Figura 66: Tempo como usuário de telefone celular.....	169
Figura 67: Freqüência de leitura e envio de SMS.....	169
Figura 68: Opinião sobre o tamanho da tela do telefone celular.....	171
Figura 69: Opinião sobre o tamanho do teclado do telefone celular.....	171
Figura 70: Opinião sobre a navegação em um telefone celular.....	172
Figura 71: Opinião sobre o tempo de conexão.....	172
Figura 72: Opinião sobre a qualidade do conteúdo.....	173
Figura 73: Opinião sobre o custo de acesso.....	174

LISTA DE QUADROS

Quadro 1: Diferenças entre o computador de mão e o computador de mesa.....	31
Quadro 2: Atributos do contexto de uso.....	61
Quadro 3: Descrição das tarefas.....	127
Quadro 4: Definição de qualidade e quantidade para as tarefas.....	130
Quadro 5: Resumo das diferenças significativas na usabilidade medida nos três contextos de avaliação.....	176
Quadro 6: Resumo da comparação dos resultados nos três contextos de avaliação.....	184

LISTA DE TABELAS

Tabela 1: Distribuição dos participantes por faixa etária e sexo.....	126
Tabela 2: Resumo das medidas e métricas de desempenho das avaliações.	180

LISTA DE SIGLAS

CDMA	Code Division Multiple Access
IHC	Interação Humano-Computador
MUSiC	Measuring Usability of Systems in Context
NASA	National Aeronautics and Space Administration
PDA	Personal Digital Assistant (Assistente Digital Pessoal).
SMS	Short Message Service
SUS	System Usability Scale
TLX	Task Load Index
WAP	Wireless Application Protocol

INTRODUÇÃO

A Interação Humano-Computador não está limitada à interação do usuário com as interfaces das páginas Web e dos programas aplicativos para o computador de mesa. O computador onipresente e invisível embarcado em diversos produtos utilizados diariamente, que as pessoas não percebem como computadores, não será mais comandado somente pelo mouse ou pelo teclado. Comandos de voz e visuais, gestos, ou até mesmo a simples presença do usuário em determinado local são novas maneiras de comandar os computadores em suas mais variadas versões (Ipiña, 2002).

É certo que a computação onipresente (*Ubiquitous Computing*) como foi vislumbrada em (Weiser,1991) ainda está mais para a realidade dos laboratórios e centros de pesquisa do que incorporada ao nosso dia-a-dia. Definições de padrões e protocolos, maiores velocidades de conexão e preços acessíveis são requisitos necessários para aplicações economicamente viáveis.

Dentro deste novo universo uma área que vem crescendo rapidamente em todo o mundo é a da Internet móvel, definida por Taurion (2002) como o uso de tecnologias de comunicação sem fio para acesso a informações e aplicações Web a partir de dispositivos móveis como telefones celulares ou *PDA's*.

A telefonia móvel juntamente com a Internet são consideradas as duas tecnologias de comunicação que mais cresceram em todos os tempos. Segundo a International Telecommunication Union o acesso à Internet através de um telefone celular deve ser a nova revolução nas comunicações (ITU,2002). A figura 1 mostra o crescimento surpreendente dessas duas tecnologias na última década.

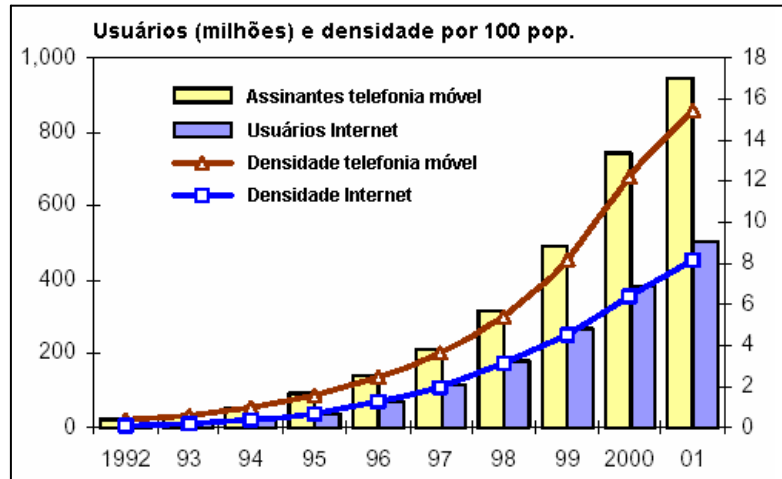


Figura 1: Número de usuários de telefonia móvel e Internet no mundo

Fonte: International Telecommunication Union (ITU, 2002)

O Brasil é considerado atualmente o maior mercado para Internet móvel na América Latina (GSA, 2002). Segundo a Agência Nacional de Telecomunicações-ANATEL em setembro de 2003 o número de assinantes de telefones celulares no Brasil chegou a 41 milhões projetando para 2005 um total de 58 milhões de assinantes como ilustra a figura 2 (Valente, 2002; ANATEL, 2000).

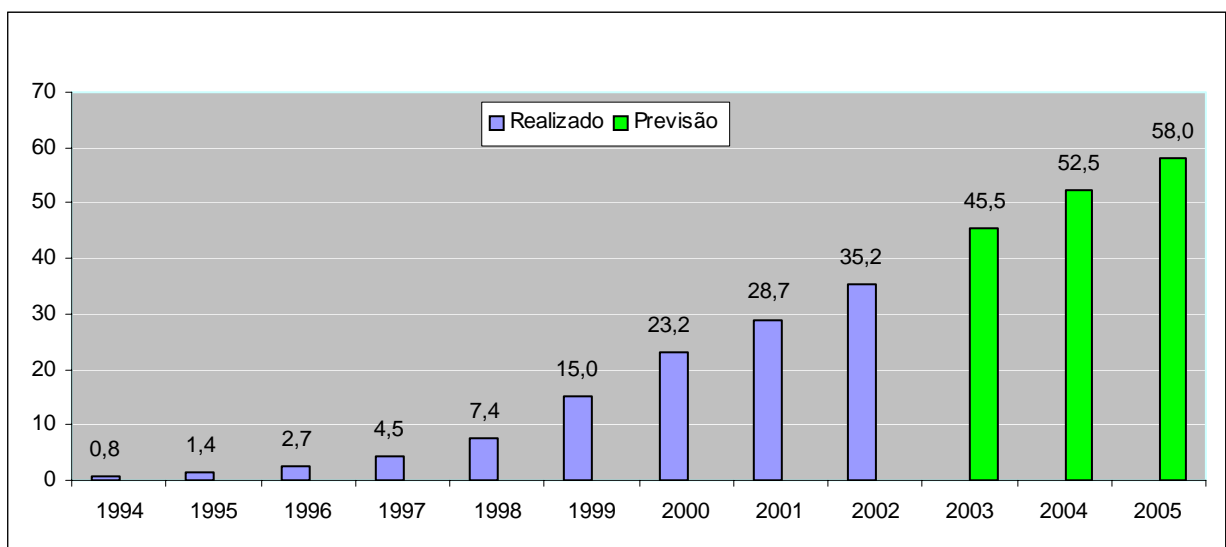


Figura 2: Planta móvel brasileira (em milhões de usuários)

Fonte: (Valente, 2002)

Em junho de 2002 o Brasil foi incluído, como único país da América Latina, na pesquisa mundial da A. T. Kearney, realizada em conjunto com a Universidade de Cambridge, para avaliar as principais tendências do uso da Internet Móvel no mundo (Mobinet #5, 2002). Esta pesquisa mostrou que 36% dos usuários de telefone celular no Brasil possuem telefone celular com acesso à Internet e que 21% já usaram algum serviço da Internet móvel. Estes números estão muito próximos da média mundial que é de 41% e 27% respectivamente. Acredita-se que em breve o número de telefones celulares supere o número de telefones fixos em serviço no País, cujo percentual hoje já representa 45% (Valente, 2002).

Este cenário aponta para uma crescente mobilidade (figura 3) a partir do aumento tanto das necessidades quanto das soluções móveis (Taurion, 2003).

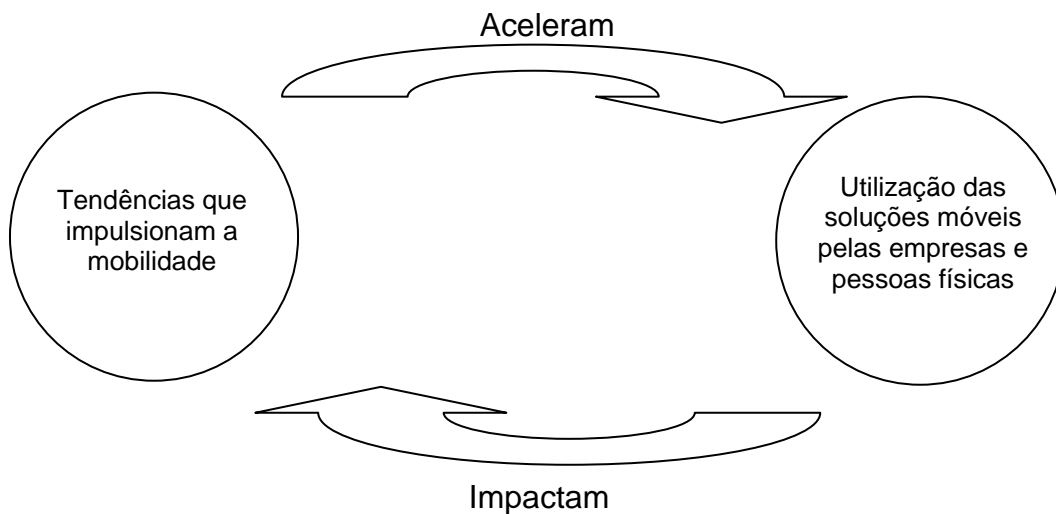


Figura 3: Fatores que promovem o incremento da mobilidade

Fonte: (Taurion, 2003)

Esse modelo implica na ocorrência de três eventos simultâneos (Taurion, 2003):

- As operadoras e os provedores de conteúdo, serviços e soluções devem oferecer o serviço e a tecnologia.
- Os computadores de mão devem ser compatíveis para suportar os serviços e a tecnologia disponíveis.

- Os usuários devem desejar, necessitar e aceitar esses serviços e tecnologias.

Segundo Taurion (2003) a aceitação do serviço tende a ser o elo mais lento desse processo. Além da utilidade, disponibilidade e custo, um dos principais elementos para a aceitação do usuário refere-se à usabilidade das interfaces e equipamentos. Segundo Chittaro (2001) a usabilidade dos equipamentos e aplicações para os computadores de mão é um fator determinante do sucesso deste segmento da computação móvel.

Desta forma, a usabilidade deve ser considerada como um dos principais requisitos de projeto para que as aplicações e serviços atendam às necessidades do usuário móvel, permitindo que ele seja eficaz, produtivo no uso do tempo e dos recursos e esteja satisfeito em relação aos atributos do sistema.

Entretanto, as técnicas tradicionais de projeto e avaliação de usabilidade que foram inicialmente desenvolvidas para os computadores de mesa precisam ser revistas e adaptadas considerando as diferenças inerentes às necessidades e características do usuário móvel, ao novo contexto de uso destes equipamentos, às características das aplicações e às suas limitações físicas como telas e teclados pequenos.

Relevância do trabalho

A seguinte declaração de Holmquist (2002) ilustra uma importante questão que se apresenta atualmente na área de Interação Humano-Computador referente à avaliação de usabilidade para os computadores de mão:

Os computadores de mão exigem que se olhe para a avaliação de usabilidade de uma maneira totalmente diferente. As avaliações tradicionais de IHC para os sistemas em computadores pessoais consistem em realizar um experimento dentro do laboratório, com o menor número possível de variáveis externas, para determinar a usabilidade do sistema.

Para os sistemas em equipamentos móveis isto simplesmente não funciona. A razão da dificuldade em se avaliar estes sistemas reside no fato de que a situação de uso destes equipamentos é completamente imprevisível, não somente em relação ao local onde o equipamento é usado mas também ao contexto social. Estas condições simplesmente não podem ser reproduzidas dentro de um laboratório e elas exercem uma enorme influência na usabilidade de um sistema.

Os pesquisadores afirmam que os critérios e métodos de IHC precisam ser reconsiderados para atender às necessidades da “interação móvel” (Johnson, 1998). Segundo Johnson (1998 apud Brewster, 2002, p.191) os métodos que ignoram o contexto de uso não são adequados para avaliar interfaces para equipamentos móveis, para o autor:

...o laboratório de usabilidade convencional não está apto a simular adequadamente aspectos tão específicos como as condições climáticas por exemplo, e não poderia reproduzir a vasta gama de atividades paralelas e necessidades dos usuários que ocorrem quando o usuário está em uma situação real de uso.

Waterson (2002) também analisa a importância do contexto de uso na avaliação de um computador de mão. Para o autor, colocar um equipamento móvel para ser testado dentro de um laboratório, normalmente “preso” sob uma câmera de vídeo, não reproduz de forma alguma o ambiente real de uso.

Os estudos citados relatam a dificuldade de reproduzir dentro do laboratório todos os elementos pertencentes a um contexto tão dinâmico e diversificado como o do usuário móvel. Entretanto, limitações de tempo e de recursos físicos e financeiros podem não permitir a realização das avaliações de usabilidade fora do laboratório e portanto a aplicação de técnicas tradicionais de avaliação de usabilidade continuam sendo utilizadas para realizar as avaliações destes equipamentos como mostram os estudos de Buchanan (2001), Vyas (2002) e Weiss (2002).

Novas soluções que permitam auxiliar na avaliação de usabilidade dos computadores de mão fora do laboratório, como equipamentos e *softwares* apropriados que possibilitem registrar a interação, ainda estão sendo pesquisadas e portanto ainda são poucas as publicações sobre este tema.

Ao contrário dos computadores de mesa, são raras as publicações que mostram estudos comparativos entre diferentes técnicas de avaliação de usabilidade para os computadores de mão como o estudo de Brewster (2002), detalhado no capítulo 3.

A partir dessas constatações é possível concluir que faltam referências na literatura que permitam avaliar concretamente qual o impacto da escolha do contexto de avaliação nos resultados da avaliação de usabilidade de um computador de mão.

Para preencher esta lacuna este trabalho se propõe a realizar uma investigação científica a partir do seguinte problema de pesquisa:

Definição do problema de pesquisa

Qual é a influência do contexto de avaliação nos resultados da avaliação de usabilidade dos computadores de mão baseada na técnica de ensaios de interação?

Hipótese

A hipótese básica que norteará o desenvolvimento desta pesquisa foi assim definida:

Na avaliação de usabilidade dos computadores de mão os ensaios de interação realizados dentro do laboratório não fornecem resultados tão satisfatórios quanto os resultados obtidos nos ensaios de interação realizados fora do laboratório, com o

usuário em uma situação mais próxima ao contexto de uso deste tipo de equipamento.

Para responder a esta questão de pesquisa este trabalho se propõe a realizar um estudo comparativo entre três diferentes abordagens para a realização de avaliação de usabilidade dos computadores de mão baseadas na técnica de ensaios de interação com os seguintes objetivos:

Objetivo Geral

- Investigar qual a influência exercida pelos diferentes contextos de avaliação nos resultados da avaliação de usabilidade da interface com o usuário de uma aplicação de Internet móvel para telefone celular, baseada na técnica de ensaios de interação.

Objetivos Específicos

- Esclarecer sobre as atividades e equipamentos necessários para a montagem e realização de ensaios de interação para avaliar a usabilidade da interface com o usuário de uma aplicação de Internet móvel para telefone celular segundo três contextos de avaliação diferentes: emulador operado em laboratório, telefone celular operado em laboratório e telefone celular operado em campo.

- Estabelecer um sistema de índices de desempenho da usabilidade e da avaliação que permita a comparação dos três contextos de avaliação.

- Conhecer os resultados da usabilidade medida em cada um dos três contextos de avaliação.
- Elaborar um guia de referência para auxiliar na escolha e na montagem de ensaios de interação para avaliar a usabilidade dos computadores de mão.

Procedimentos metodológicos

A pesquisa realizada pode ser classificada como aplicada do ponto de vista de sua natureza, pois teve como objetivo gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos (Gil, 1991).

Inicialmente foi realizado um estudo exploratório a partir de uma extensa pesquisa bibliográfica que teve como objetivos determinar o estado da arte na área de Interação-Humano Computador para os computadores de mão.

Foi realizado um extenso levantamento bibliográfico buscando informações sobre as principais características físicas dos computadores de mão, o contexto de uso destes equipamentos e o perfil do usuário móvel. A pesquisa procurou identificar o estado da arte na área de avaliação de usabilidade para os computadores de mão a partir das informações sobre como essas avaliações vêm sendo conduzidas.

Na seqüência, diversas fontes foram consultadas buscando identificar as várias definições e técnicas de avaliação de usabilidade, com o objetivo de estabelecer a usabilidade voltada aos computadores de mão dentro da perspectiva da Engenharia de Usabilidade.

Esta pesquisa foi realizada em anais de congressos, revistas científicas, bases de dados eletrônicas, livros, dissertações e teses com o objetivo de aprofundar o conhecimento sobre o tema da Interação Humano-Computador para os

computadores de mão e, particularmente, sobre as avaliações de usabilidade destes equipamentos.

Fez-se necessária também uma extensa pesquisa documental, em publicações administrativas da ANATEL e de outros órgãos internacionais, para levantar os principais índices estatísticos que permitiram identificar um expressivo crescimento do uso dos computadores de mão no mundo inteiro e, particularmente, no Brasil. Estas fontes forneceram indicadores importantes não só sobre os equipamentos, mas também sobre o comportamento do usuário móvel em relação à utilização dos serviços e aplicações disponíveis para os computadores de mão.

Uma vez definida a avaliação de usabilidade dos computadores de mão como objeto deste estudo, e os diferentes componentes dos contextos de avaliação como as variáveis que são capazes de influenciar nos resultados das avaliações, para testar a hipótese levantada foi realizada uma pesquisa experimental de abordagem quantitativa e qualitativa.

A partir da pesquisa bibliográfica e documental foi possível definir as seguintes etapas para esta pesquisa experimental:

- Definição dos contextos de avaliação para a realização dos ensaios de interação, incluindo equipamentos, condições da avaliação e critérios para escolha da amostra de participantes, tarefas e medidas e métricas de usabilidade.
- Definição dos índices de desempenho para comparar os resultados obtidos nos três contextos de avaliação.
- Realização dos ensaios de interação segundo os três contextos de avaliação definidos.
- Comparação dos resultados da usabilidade medida e dos índices de desempenho das avaliações nos três contextos de avaliação definidos.

Todas estas etapas da pesquisa experimental, com os seus objetivos, e respectivos métodos utilizados para alcançá-los, estão descritas detalhadamente no capítulo 4.

Limitações da pesquisa

Entre os diversos computadores de mão e aplicações existentes, foi escolhido para ser avaliado nesta pesquisa um portal de Internet móvel que pode ser acessado via telefone celular.

Entre as várias possíveis definições para os contextos de avaliação foram escolhidas, a partir da revisão bibliográfica, três abordagens principais. Os dois contextos de avaliação definidos dentro do laboratório fazem uso de métodos amplamente utilizados nos ensaios de interação para os computadores de mão: os emuladores e as câmeras de documentos. O terceiro contexto de avaliação apresenta uma proposta para a realização dos ensaios de interação fora do laboratório na qual, a partir do desenvolvimento de uma minicâmera sem fio, foi possível colocar o usuário em uma entre as diversas possíveis situações de uso de um computador de mão.

Acredita-se que a escolha de outros modelos de computadores de mão e aplicações, bem como a definição de novos contextos de avaliação, se constituem em importantes temas para a realização de trabalhos futuros.

Estrutura do trabalho

O Capítulo 1 apresenta os computadores de mão mostrando as principais características físicas desses equipamentos, o contexto de uso e o perfil do usuário móvel.

O Capítulo 2 define usabilidade especificando seus principais componentes e discorre sobre as técnicas de usabilidade, especialmente ensaios de interação.

O Capítulo 3 apresenta as principais abordagens que estão sendo utilizadas para realização das avaliações de usabilidade para os computadores de mão.

O Capítulo 4 descreve os procedimentos metodológicos utilizados para a realização dos ensaios de interação baseados em três diferentes contextos de avaliação.

O Capítulo 5 apresenta e analisa os resultados obtidos nas avaliações de usabilidade.

O Capítulo 6 é dedicado à interpretação e à discussão dos resultados, à conclusão e a sugestões de trabalhos futuros.

1 COMPUTADORES DE MÃO

O avanço da tecnologia e o aumento da capacidade de processamento dos componentes eletrônicos vêm permitindo o surgimento de novas concepções do computador pessoal portátil. Esta portabilidade aliada às novas tecnologias de acesso a redes sem fio se traduz em um conceito fundamental para a área de tecnologia de informação e comunicação: a mobilidade do usuário.

A Tecnologia da Informação e a Internet quebraram a barreira do tempo, a mobilidade quebra também a barreira do espaço tornando possível, a qualquer hora e em qualquer lugar, o acesso a informações e serviços que antes só eram obtidos através do computador pessoal que estava fixo sobre a mesa, em casa ou no local de trabalho (Taurion, 2003).

Weiss (2002) classifica o computador pessoal em quatro categorias principais conforme a figura 4.

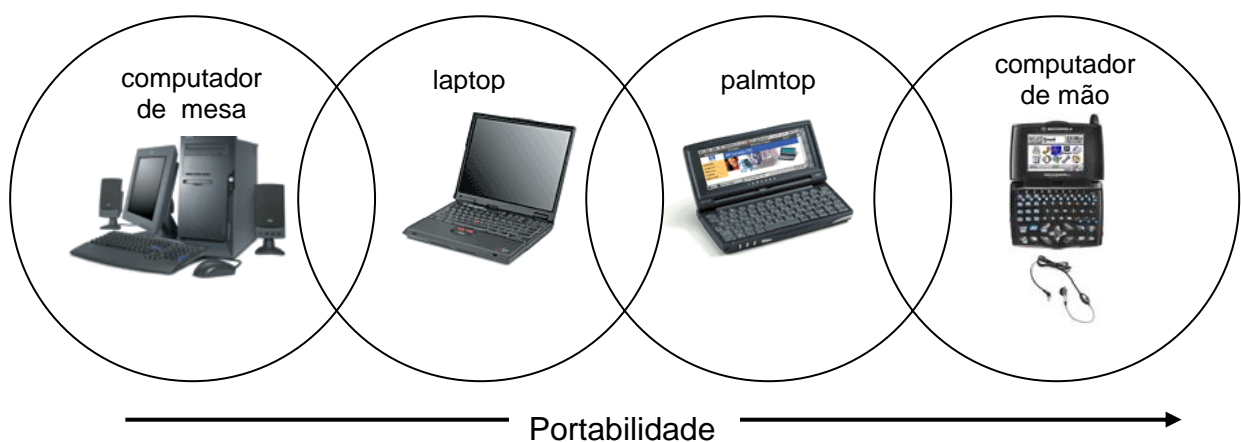


Figura 4: O computador pessoal

Fonte: (Weiss, 2002)

Os *laptops* são os primeiros computadores pessoais portáteis, com dispositivos de entrada e saída de dados muito similares ao computador de mesa. Os *palmtops* são menores que os *laptops* e já podem ser levados dentro do bolso ou da bolsa, entretanto ainda são melhor operados quando apoiados sobre uma superfície. Para ser considerado um computador de mão o equipamento deve atender às três seguintes características (Weiss, 2002):

- Deve funcionar sem cabos, exceto temporariamente, como para a recarga elétrica ou para sincronizações com o computador de mesa.
- Deve ser operado facilmente com uma única mão, sem a necessidade de estar apoiado sobre uma mesa.
- Deve permitir a adição de novos aplicativos ou suportar conexão à Internet.

De acordo com essa definição podem ser considerados exemplos de computadores de mão os PDAs (*Personal Digital Assistants*), os telefones celulares e os equipamentos híbridos que acumulam as características dos dois anteriores.

1.1 Características dos computadores de mão

São várias as diferenças entre o computador de mesa e os computadores de mão em relação ao equipamento, à finalidade e ao contexto de uso como ilustra o quadro 1.

O computador de mesa é fixo, está sempre conectado via cabos, possui memória e capacidade de armazenamento praticamente ilimitadas, tela grande e colorida e entrada de dados via mouse ou teclado.

Quadro 1: Diferenças entre o computador de mão e o computador de mesa

	Computador de mão	Computador de mesa
Uso	Consulta rápida à informação em várias situações diferentes, normalmente longe do computador de casa ou do escritório. Comunicação rápida por voz e/ou mensagem de texto.	Longas e complexas tarefas de processamento da informação. <i>Web browsing</i> e <i>e-mail</i> .
Freqüência/Duração	Várias vezes durante o dia com curta duração.	Poucas vezes durante o dia durante longo período de tempo.
Mobilidade	Pode ser facilmente operado com as mãos sem necessidade de apoio. Pode ser levado no bolso com facilidade.	Necessita de uma mesa. Precisa estar conectado por cabos ou tem necessidade de recarga freqüente. Alguns modelos são portáteis.
Entrada de dados	Diversas: Minitel, caneta, botões de navegação.	Padrão: Teclado completo, mouse, tela de toque.
Display	640X480 <i>pixels</i> para menos.	640X480 <i>pixels</i> para mais.
Capacidade de Armazenamento	Normalmente pouca. Alguns equipamentos possuem expansão de memória.	Praticamente ilimitada. Disquetes, CDRW, DVD RAM, disco rígido.
Conexão	Lenta e instável (por enquanto).	Veloz e estável.

Fonte: Adaptado de (Weiss, 2002) e (Chincholle, 2002a)

Apesar de serem mais leves, portáteis e permitirem conexão sem fio, os computadores de mão têm pouca memória, pouca ou quase nenhuma capacidade de armazenamento, bateria que deve ser sempre recarregada, conexão ainda lenta e muitas vezes pouco confiável.

Limitações como velocidade de conexão, duração da bateria e capacidade de armazenamento podem vir a desaparecer, ou pelo menos, ser minimizadas à medida que a tecnologia evoluir. No entanto dois limitadores principais dificilmente serão alterados devido à própria característica do equipamento: a visualização em telas pequenas e a entrada de dados em espaço reduzido, que serão discutidos com mais detalhes a seguir.

1.1.1 Telas pequenas

Todos os computadores de mão apresentam uma característica em comum: as telas pequenas. A figura 5 mostra alguns exemplos de telas de computadores de mão.



Figura 5: Exemplos de telas pequenas: PDA e telefone celular

Fonte: (Palm, 2003)
© Copyright 2003. Palm Inc
Todos os direitos reservados.

Fonte: (Orange, 2003)
© Copyright 2003 Orange.
Todos os direitos reservados.

As telas pequenas tornam a leitura mais difícil e exigem novos modelos de apresentação de conteúdo para maximização de espaço sem sobrecarregar a área de visualização. São diversos os modelos de telas que variam quanto à dimensão, ao uso de cores, a ser sensível ao toque e à possibilidade de mostrar *bitmaps*.

As telas devem continuar pequenas e podem até ser menores que as dos telefones celulares atuais. A figura 6 mostra um relógio com vários recursos de PDA como agenda de compromissos, catálogo de telefones, calendário e calculadora. A tela é sensível ao toque e pode ser ativada com uma mini-caneta que está embutida na pulseira (Fossil, 2003).



Figura 6: Exemplo de tela pequena: relógio de pulso

Fonte: (Fossil, 2003)

© Copyright 2003 Fossil, Inc. Todos os direitos reservados.

Neste exemplo de computador de mão o usuário visualizará, em pequenas telas coloridas, as horas em função da sua localização, informações sobre o tempo, conteúdo da Internet e mensagens de texto, além de poder armazenar e reproduzir músicas, vídeos e fotografias (Hansell, 2003; Microsoft, 2003a).

1.1.2 Entrada de dados

Quanto à entrada de dados, não existe ainda para os computadores de mão uma tecnologia padronizada como o teclado e o mouse para o computador de mesa. A forma de entrada de dados difere de um equipamento para outro, podendo ser na forma de botões e controles para facilitar a navegação, miniteclados ou tela de toque.

Os telefones celulares normalmente possuem miniteclados alfanuméricos nos quais é preciso pressionar várias vezes a mesma tecla até encontrar a letra desejada. Este é um método conhecido como *multi-tap* que, apesar de ser um método de entrada de dados padrão em todo o telefone celular, é considerado difícil

e pouco intuitivo (Weiss, 2002). Uma alternativa para a entrada de dados, mantendo o mesmo teclado, é o uso de texto preditivo, ou seja, o usuário vai pressionando as teclas que contém a letra desejada uma única vez e o *software* vai adivinhar o que ele está querendo escrever a partir de um banco de palavras. Alguns estudos de usabilidade (Weiss, 2001) verificaram que o usuário fica confuso em função da diferença entre o que ele está querendo digitar e o que está aparecendo na tela. Esta tecnologia conhecida como T9[®] (T9, 2003) está presente em vários modelos de telefones celulares como um método alternativo para a entrada de dados.

Outras alternativas são o uso do teclado Fastap[™] (Digitwireless, 2003) que é um teclado com leves saliências onde estão as letras, e depressões onde estão os números (figura 7), desta forma cada caracter pode ser pressionado diretamente, uma única vez. Este teclado também possui uma versão QWERTY em tamanho menor ao de um cartão de crédito.



Figura 7: Teclado Fastap[™]

Fonte: (Digitwireless, 2003)
© Copyright 2003 Digitwireless. Todos os direitos reservados.

Os miniteclados QWERTY, comuns em PDA's, também estão presentes em alguns modelos de telefones celulares (figura 8).



Figura 8: Minitclado QWERTY em PDA e telefone celular

Fonte: (Handspring, 2003)
 © Copyright 1999-2003 Handspring.
 Todos os direitos reservados.

Fonte: (Nokia, 2003)
 © Copyright 2003 Nokia Corporation.
 Todos os direitos reservados.

Muitos modelos de computadores de mão, inclusive telefones celulares (figura 9), incorporaram os princípios da manipulação direta permitindo que o usuário manipule objetos da tela usando algum tipo de apontador. Desta forma é possível utilizar uma caneta para selecionar as teclas virtuais que aparecem na tela ou escrever diretamente sobre ela quando o equipamento possuir *softwares* de reconhecimento de escrita.



Figura 9: Telefone celular com tela de toque

Fonte: (Sony Ericsson, 2003)
 © Copyright 2001-2003. Sony Ericsson Mobile Communications AB.
 Todos os direitos reservados.

São várias as pesquisas, como os estudos de Chincholle (1999) e Silfverberg (2003), que buscam encontrar novas soluções para a entrada de dados nos computadores de mão, este que é considerado um dos maiores problemas de

interação com os equipamentos pequenos. Algumas idéias são curiosas como um teclado virtual que é a projeção de raios laser sobre uma superfície plana (figura 10). Ao detectar quais linhas foram interrompidas o dispositivo sabe quais teclas foram ativadas (VKB, 2003).



Figura 10: Teclado virtual por infravermelho

Fonte: (VKB, 2002)

© Copyright 2002. VKB. Todos os direitos reservados.

Há também algumas tecnologias que utilizam sensores colocados nas mãos (figura 11) que detectam os movimentos dos dedos do usuário mapeando esses movimentos com a tecla que ele teve a intenção de pressionar (Chincholle, 1999; Senseboard, 2002; Kittytech, 2002).



Figura 11: Teclados virtuais com sensores

Fonte: (Senseboard, 2002)

© Copyright 2002 Senseboard Technologies AB.
Todos os direitos reservados.

Fonte: (Kittytech, 2002)

© Copyright 2002 Kittytech Project.
Todos os direitos reservados.

O trabalho de Mackenzie (2002) apresenta uma extensa e detalhada pesquisa sobre os principais meios de entrada de dados em computadores de mão e (Little Springs Design, 2002) é uma fonte de referências atualizadas para as principais novidades tecnológicas nessa área.

1.2 O usuário móvel e o contexto

Além das características físicas dos computadores de mão, é muito importante entender o usuário móvel e o contexto em que ele está inserido.

Enquanto o computador de mesa é usado para digitar texto, elaborar planilhas eletrônicas, fazer pesquisas na Web, ou seja, tarefas que exigem concentração e são executadas durante um longo período de tempo, os computadores de mão são voltados para aplicações mais rápidas. O tempo é um fator muito importante para o usuário móvel. Ele é mais impaciente e exigente que o usuário do computador de mesa e tende a utilizar serviços que permitam uma manipulação rápida da interface e um acesso à informação através de um número reduzido de passos (Ericsson, 2001).

O ambiente de uso de um computador de mão é normalmente muito mais dinâmico em relação ao ambiente de uso do computador de mesa e, portanto, demanda a atenção do usuário para outras tarefas além da interação com o equipamento. O usuário de um computador de mão normalmente está envolvido em várias atividades que ocorrem simultaneamente. A sua atenção está dividida entre o uso do equipamento, as outras atividades que ele está realizando e o ambiente que o cerca (Chincholle, 2002b). Desta forma, o foco principal não é a interação com o equipamento, mas a busca da informação enquanto ele interage com o ambiente à

sua volta. Segundo Rischpater (2000, p.33) “o usuário móvel tem menor capacidade de processar e absorver conteúdo que um usuário que está sentado em frente a um computador de mesa”.

O usuário móvel pode estar lendo uma mensagem de texto que acabou de receber em seu telefone celular enquanto caminha pela rua, pode estar parado em uma longa fila de check-in no aeroporto e aproveitar o tempo para atualizar sua agenda de compromissos no PDA enquanto conversa com sua secretária pelo telefone celular, ou pode estar sentado em um restaurante procurando a próxima sessão do cinema enquanto janta e conversa com amigos. Em meio a tantas tarefas, a interação com o equipamento deve exigir o menor nível possível de concentração por parte do usuário. Para Chincholle (2002b) a interação com o equipamento é uma tarefa secundária que não deve interferir com a tarefa principal que ele está realizando, como por exemplo dirigir ou caminhar. Uma pesquisa realizada em Londres para avaliar um computador de mão com acesso a páginas Web constatou que a interação com o equipamento exigia uma grande atenção e concentração do usuário dificultando a sua percepção sobre o que estava acontecendo à sua volta. Os usuários relataram que se sentiam inseguros ao utilizar o equipamento enquanto estavam na rua, principalmente à noite, pois se achavam alvos fáceis para possíveis ladrões (Helyar, 2002).

Os computadores de mão estão destinados a atingir uma grande variedade de usuários, entre eles estão aqueles que nunca utilizaram o computador, ou que têm pouca experiência no seu uso. A “usabilidade imediata” é muito importante para esses usuários (Mackenzie, 2002). Neste contexto, fazer uso de modelos e metáforas de interface com o usuário que são amplamente utilizadas no computador de mesa pode não ser apropriado. Além de não estar familiarizado com esses

elementos, o usuário pode considerar o equipamento muito complicado e não estar disposto a aprender formas complexas de interação (Mohageg, 2000).

As necessidades e características do usuário móvel, o contexto de uso dos computadores de mão, as características das aplicações e as limitações físicas desses equipamentos são fatores que irão influenciar a interação e devem ser considerados tanto no projeto das interfaces quanto na avaliação de usabilidade destes equipamentos.

2 USABILIDADE

A Interação Humano-Computador (IHC) pode ser definida como uma área do conhecimento relacionada ao projeto, à avaliação e à implementação de sistemas computacionais interativos para uso humano e ao estudo dos principais fenômenos que os cercam (Hewett, 1996).

A IHC tem suas raízes nas áreas de ergonomia, psicologia, sociologia, antropologia, engenharias, ciência da computação e desenho industrial. É esse referencial teórico que vai fornecer importantes subsídios para compreender quais os processos que estão envolvidos na interação física e na interação cognitiva do ser humano com todo o sistema computacional.

Esse processo de interação se dá através da interface com o usuário que pode ser definida como o conjunto de todas as linguagens através das quais o usuário e o produto se comunicam (Mayhew, 1999). Essas linguagens de comunicação se apresentam nas mais variadas formas como os elementos visuais das telas (p.ex. menus, ícones, cores e links), os componentes físicos (p.ex. mouses, teclados e botões), os comandos visuais e de voz, gestos, até as novas gerações de interfaces pertencentes aos computadores que estarão espalhados por todo o ambiente, dotados de sensores capazes de captar informações sobre o contexto e reagir às pessoas e às atividades que estiverem à sua volta (Ipiña, 2002).

Desta forma, a interface com o usuário em qualquer sistema interativo tem papel fundamental, pois é ela que possibilita a comunicação entre o usuário e o sistema, ou seja, quanto maior for a usabilidade da interface, mais fácil será a comunicação.

A usabilidade passa então a ser um objetivo a ser atingido durante todo o projeto da interface com o usuário.

A usabilidade, que se tornou um fator importante no desenvolvimento de software e vem aumentando a sua importância também no desenvolvimento de outros produtos, pode ser considerada como um conjunto de métodos na Engenharia de Usabilidade ou como uma filosofia de desenvolvimento de projetos no Projeto Centrado no Usuário (Keinonen, 1998).

A Engenharia de Usabilidade é definida como o processo que vai fornecer os métodos estruturados para que se possa atingir um alto grau de usabilidade da interface com o usuário durante o desenvolvimento do produto (Mayhew, 1999). O Projeto Centrado no Usuário tem como objetivo produzir sistemas fáceis de aprender e usar, seguros e efetivos em facilitar as atividades do usuário (Rocha, 2003). É uma filosofia que coloca o usuário no centro de todo o processo, ou seja, os objetivos do produto, seu contexto de uso e todos os aspectos das tarefas que o produto irá suportar são derivados a partir da perspectiva do usuário. Alguns autores (Jokela, 2003; Rubin, 1994) consideram que são dois termos para um mesmo conceito pois os processos de Projeto Centrado no Usuário e da Engenharia de Usabilidade são similares uma vez que ambos têm como objetivo estabelecer procedimentos e padrões para o envolvimento efetivo do usuário ao longo de todo o processo de desenvolvimento do projeto de sistemas interativos (Öörni, 2003).

As técnicas de usabilidade podem e devem ser aplicadas em várias fases durante o ciclo de desenvolvimento do produto, desde técnicas aplicadas nas fases iniciais do projeto até testes de usabilidade realizados com o produto final em campo, cujos resultados irão subsidiar o desenvolvimento de novas versões ou até mesmos novos produtos.

A usabilidade como filosofia que influencia o processo de projeto é transferida ao usuário na forma de atributos concretos do produto podendo desta forma ser medida a partir da interação do usuário com o sistema e da sua experiência pessoal em relação ao seu uso (Keinonen, 1998).

2.1 Definições para Usabilidade

2.1.1 Usabilidade segundo Shackel

Na visão de Shackel (1991) a usabilidade é uma das componentes da aceitabilidade de um produto, que é o nível mais alto do modelo de percepção do usuário em relação ao produto (figura 12).

Para Shackel a utilidade se refere à combinação entre as necessidades do usuário e as funcionalidades do produto, enquanto que a usabilidade se refere à habilidade do usuário em poder utilizar essas funcionalidades na prática. A agradabilidade se refere à avaliação afetiva por parte do usuário e o custo inclui não só o custo financeiro mas também conseqüências organizacionais.

O autor sugere alguns critérios para avaliar a usabilidade em função dos seguintes componentes:

- **Eficiência:** avalia o resultado da interação em função da velocidade e dos erros.
- **Aprendizado:** avalia a relação da performance ao tempo de treinamento e à frequência de uso.
- **Flexibilidade:** avalia a adaptação a tarefas e ambientes além daqueles especificados inicialmente.

- Atitude: avalia os níveis de desgaste do usuário em termos de cansaço, desconforto, frustração e esforço pessoal.

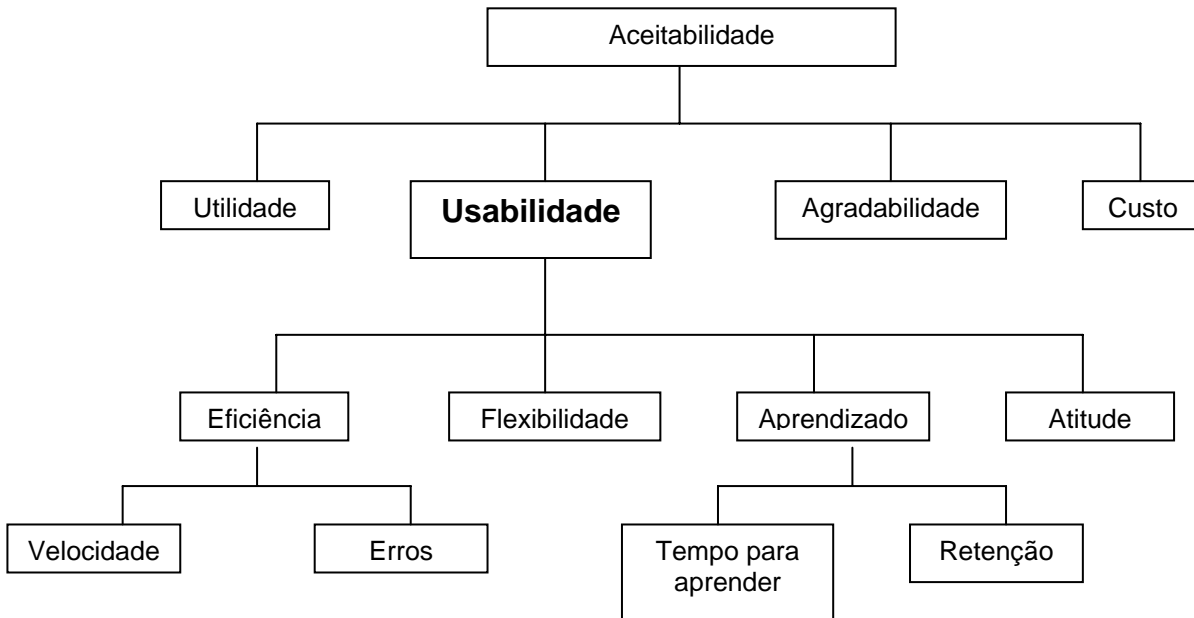


Figura 12: Modelo de atributos de aceitabilidade de um sistema segundo Shackel

Fonte: (Shackel, 1991)

Para Shackel a usabilidade é uma propriedade do sistema que não é constante, mas sim relativa aos usuários, à sua experiência, às tarefas e ao ambiente de uso.

2.1.2 Usabilidade segundo Nielsen

Nielsen (1993a) também considera a usabilidade como um entre os vários aspectos que podem influenciar a aceitabilidade de um produto. Ou seja, ela faz parte de um conceito mais amplo que envolve aspectos como custo, confiabilidade e até mesmo aceitabilidade social (figura 13). Para o autor a aceitabilidade de um sistema se refere à capacidade do sistema de satisfazer todas as necessidades e

exigências dos usuários, que podem ser tanto os usuários finais quanto outras pessoas que estejam envolvidas de alguma forma com esse sistema.

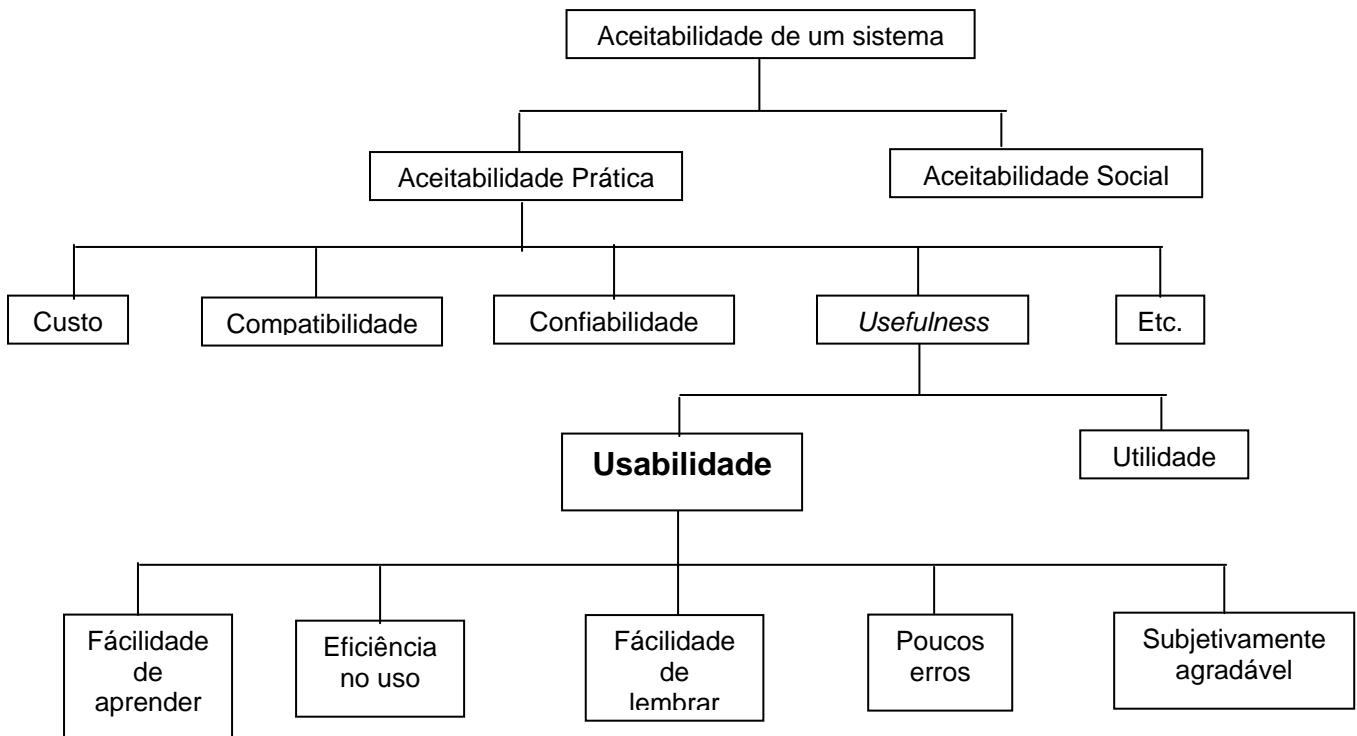


Figura 13: Modelo de atributos de aceitabilidade de um sistema segundo Nielsen

Fonte: (Nielsen, 1993a)

O termo *usefulness* se refere ao sistema poder ser usado para atingir um determinado objetivo e é decomposto em utilidade e usabilidade. A utilidade se refere a verificar se o sistema tem as funcionalidades certas e se estas funcionalidades fazem o que deve ser feito enquanto que a usabilidade se refere à qualidade com que os usuários conseguem utilizar essas funcionalidades. A usabilidade se aplica a todos os aspectos do sistema com os quais a pessoa pode interagir, incluindo os procedimentos de instalação e manutenção e, principalmente, deve ser sempre medida relativamente a determinados usuários executando determinadas tarefas.

Para que a usabilidade possa ser avaliada e medida Nielsen (1993a) a define em função destes cinco atributos:

- **Aprendizagem:** o sistema deve ser de fácil aprendizado para que o usuário possa começar a utilizá-lo rapidamente.
- **Eficiência:** o sistema deve ser eficiente no sentido de que uma vez que o usuário aprenda a utilizá-lo ele o faça com alta produtividade.
- **Memorização:** o sistema deve ser de fácil lembrança, ou seja, ao passar um determinado período sem utilizar o sistema o usuário pode utilizá-lo novamente sem ter que aprender tudo novamente.
- **Erros:** a taxa de erros deve ser baixa. Erros de extrema gravidade não devem ocorrer. Ao cometer algum erro, o usuário deve ter a possibilidade de recuperar o sistema para o estado imediatamente anterior ao erro.
- **Satisfação:** os usuários devem gostar do sistema. Ele deve ser agradável de ser utilizado para que as pessoas se sintam satisfeitas com o seu uso.

2.1.3 Usabilidade segundo Jordan

Jordan (1998) apresenta uma visão de usabilidade baseada inicialmente em três componentes: intuitividade, aprendizagem e performance do usuário com experiência (Jordan et al., 1993 apud Jordan, 1998) e que mais tarde teve uma extensão com a adição de mais dois componentes: potencial do sistema e re-usabilidade (Jordan, 1994 apud Jordan, 1998). Cada um desses componentes é definido como:

- **Intuitividade:** é o custo para o usuário (em termos de tempo e taxas de erros, por exemplo) para utilizar um produto ou realizar uma nova tarefa pela primeira vez.

- Aprendizagem: é o custo para o usuário em atingir um determinado nível de competência na realização de uma tarefa, excluindo as dificuldades encontradas para realizá-la pela primeira vez.
- Performance do usuário experiente (PUE): é o nível da performance atingido por determinado usuário ao realizar muitas vezes determinadas tarefas com um determinado produto.
- Potencial do sistema: representa o nível máximo de performance que pode ser atingido ao realizar uma determinada tarefa com um produto. É o limite máximo do PUE.
- Re-usabilidade: indica uma possível diminuição da performance que pode ocorrer após o usuário não utilizar o produto, ou não executar uma determinada tarefa, por um determinado período de tempo.

Para Jordan (1998) a usabilidade também não pode ser considerada isoladamente pois ela é uma propriedade da interação entre o produto, o usuário, a tarefa que ele está tentando executar e o ambiente que o cerca.

2.1.4 Usabilidade e as normas

2.1.4.1 ISO 9241-11

A norma ISO 9241, de 1998, sobre “Requisitos ergonômicos para trabalho de escritórios com computadores” contém 17 partes que abordam diferentes aspectos referentes ao ambiente de trabalho e a práticas do projeto de diálogo utilizado (NBR 9241-11, 2002).

A parte 11 desta norma (ISO 9241-11, 1998) apresenta as “Orientações sobre usabilidade” e define usabilidade como sendo: “a eficiência, a eficácia e a satisfação com as quais determinados usuários realizam determinadas tarefas em um determinado contexto de uso”. Onde eficácia é a exatidão e completude com as quais os usuários atingem objetivos específicos. Eficiência são os recursos gastos em relação à exatidão e completude com as quais os usuários atingem os objetivos. Satisfação é o conforto e a aceitabilidade no uso. E o contexto de uso consiste de usuários, tarefas, equipamentos (*hardware, software* e materiais), e os ambientes físico e social em que o produto é utilizado (NBR 9241-11, 2002).

Esta definição ressalta que a usabilidade não é uma propriedade isolada do produto, mas que depende de quem o está utilizando, com quais objetivos e em que contexto de uso ou ambiente o produto está sendo utilizado.

A ISO 9241-11 explica como identificar as informações que são necessárias para especificar ou avaliar a usabilidade em termos de medidas da performance do usuário e de sua satisfação. São dadas orientações, na forma de princípios e técnicas, sobre como descrever explicitamente o contexto de uso do produto e as medidas relevantes de usabilidade.

2.1.4.2 ISO/IEC 9126

A norma ISO/IEC 9126-1 (2001) “Engenharia de software: Qualidade do produto”, é uma norma para avaliação de qualidade de produto de software.

Esta norma traz o conceito de “qualidade em uso” que é definida como “a capacidade do produto de software de permitir a usuários específicos atingir metas

especificadas com eficácia, produtividade, segurança e satisfação em um contexto de uso especificado” (ABNT, 1999).

Esta definição é similar à definição de usabilidade da ISO 9241-11, e também faz referência ao contexto de uso, ressaltando que a qualidade em uso não é uma característica intrínseca do produto.

A qualidade em uso é a visão do usuário sobre a qualidade de um sistema e é medida em termos de resultados de utilização do software e não em função de suas propriedades (Bevan, 2002). Essa visão do modelo de qualidade levando em conta os resultados a partir do uso do software no ambiente especificado é uma inovação em relação ao modelo original da ISO 9126 de 1991 (ABNT,1999).

A ISO/IEC 9126-1 descreve um modelo de qualidade que categoriza os atributos de qualidade de software em seis características: funcionalidade, confiabilidade, usabilidade, eficiência, manutenibilidade e portabilidade onde cada uma dessas características é subdividida em subcaracterísticas como ilustra a figura 14.

Neste modelo a usabilidade é uma das componentes da qualidade de software e é definida como: “a capacidade do software de ser compreendido, de ser aprendido, de ser operado e de ser atraente para o usuário, quando utilizado sob condições específicas”. É uma definição mais próxima à visão de Nielsen e Jordan onde a usabilidade está mais relacionada à facilidade de uso e aprendizado.

Segundo Bevan (1999b) “atingir a qualidade em uso depende de atingir os critérios para as medidas externas de comportamento do software, o que por sua vez, depende de atingir os critérios relacionados para as medidas internas”, portanto os atributos dos três níveis devem ser medidos.

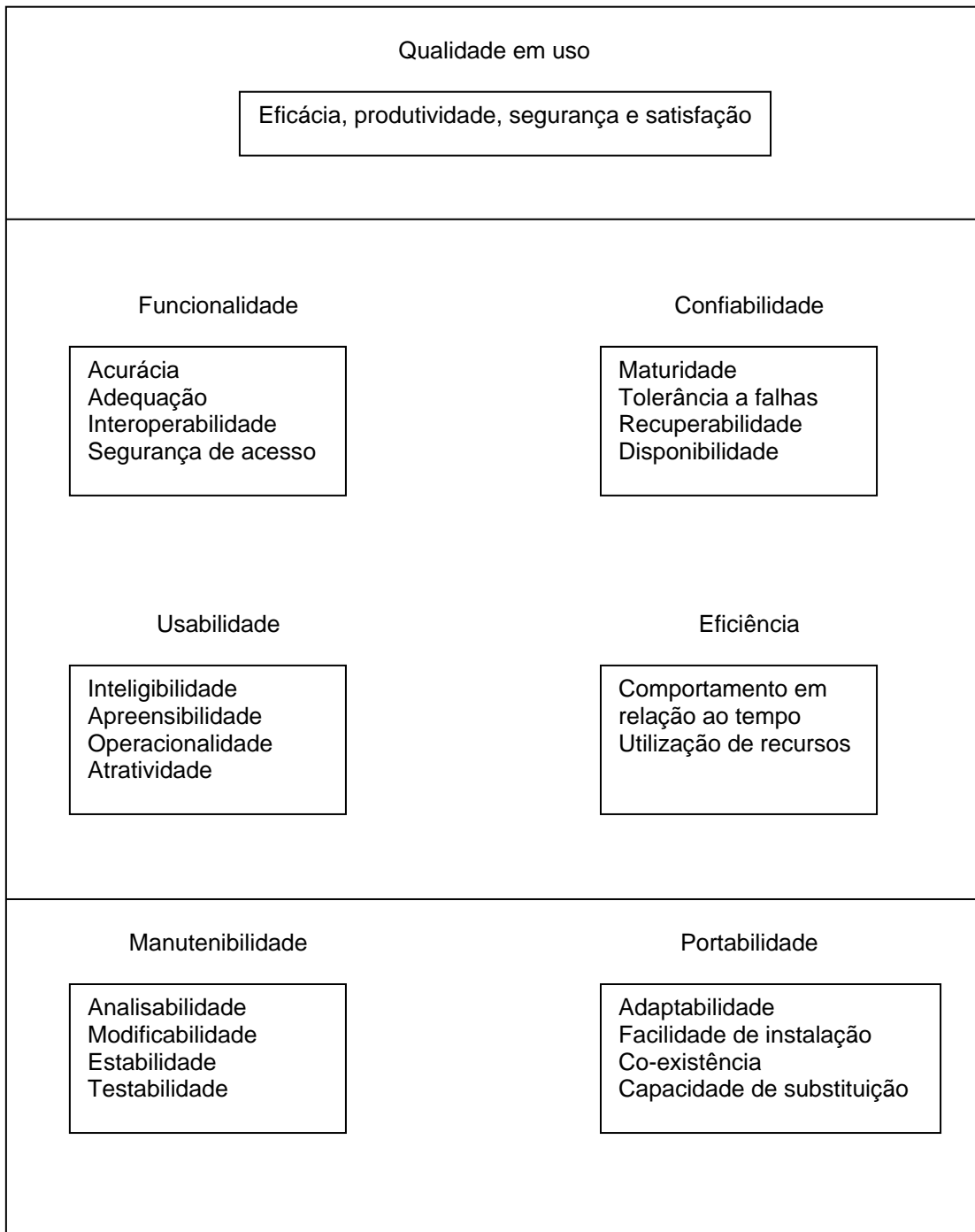


Figura 14: Modelo de qualidade para um produto de software ISO/IEC 9126-1

Fonte: (Bevan, 2001; ABNT, 1999)

Os documentos ISO/IEC 9126-2 e ISO/IEC 9126-3 definem exemplos de métricas externas e internas que se associam aos atributos de qualidade relacionados às características e subcaracterísticas deste modelo. A parte 2, métricas externas, descreve métricas que são aplicadas ao produto de software executável, enquanto a parte 3, métricas internas, descreve métricas que são aplicadas ao produto de software não-executável, durante as fases de codificação. As métricas para avaliar a qualidade em uso estão descritas no documento ISO/IEC 9126-4 (ABNT,1999).

2.1.4.3 ISO/IEC 14598

A norma ISO/IEC 14598 (partes 1 a 6) descreve o processo a ser utilizado para a avaliação de um produto de software. Ela está organizada em diferentes objetivos de avaliação segundo o ponto de vista do desenvolvedor, do adquirente e do avaliador independente.

A ISO/IEC 14598-1 (1999) também define o conceito de qualidade em uso como “a extensão na qual um produto utilizado por usuários específicos atende às suas necessidades de atingir objetivos específicos com eficácia, produtividade e satisfação em um contexto de uso específico” (Bevan, 2002). Desta forma a qualidade em uso é a visão de qualidade a partir do ponto de vista do usuário e é medida em função dos resultados da utilização do sistema e não das suas propriedades (Bevanb, 1999).

Esta norma pode utilizar qualquer modelo de qualidade, no entanto a aplicação deste processo de avaliação é muito mais simples se for utilizado o modelo da 9126-1, pois todas as normas da família 14598 estão fortemente relacionadas a este modelo de qualidade (ABNT,1999).

A figura 15 ilustra o processo de avaliação de produto de software baseado na ISO/IEC 14598-1.

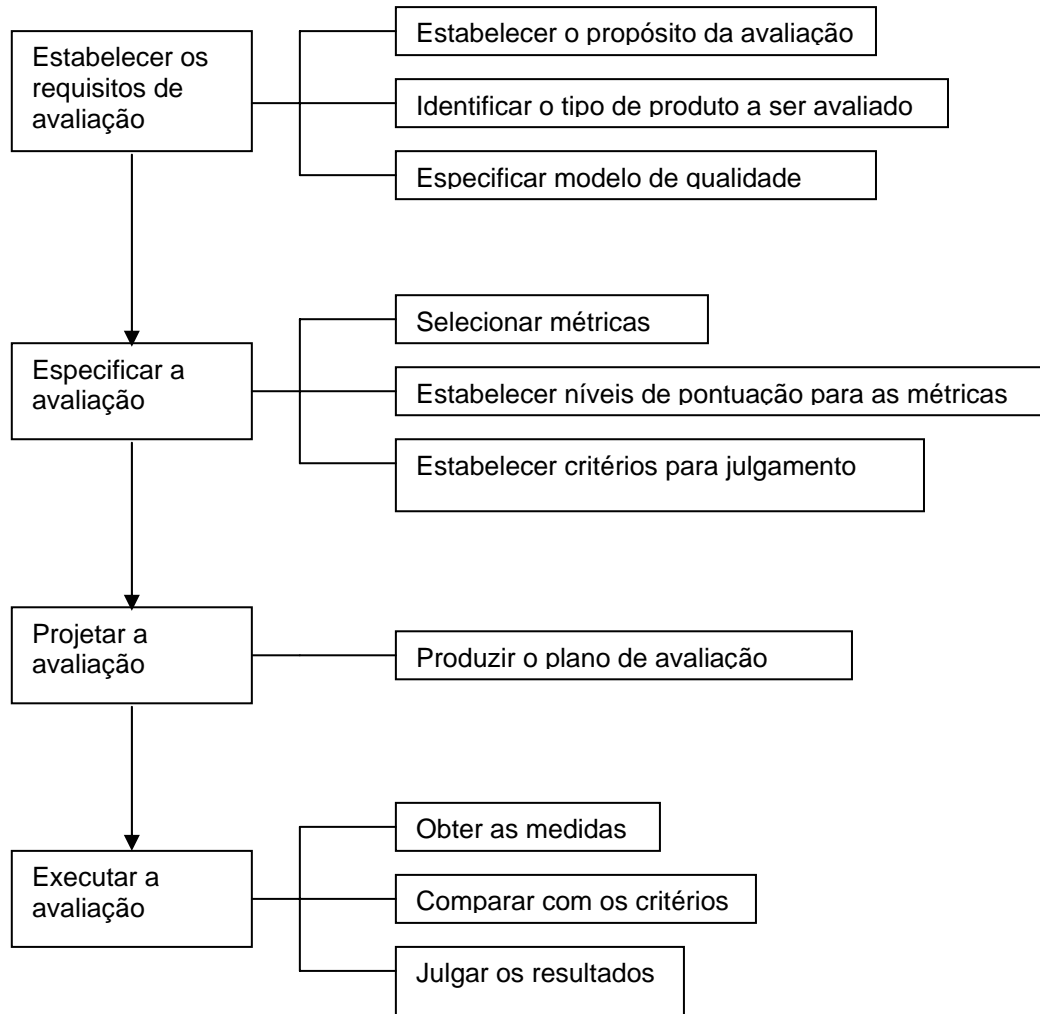


Figura 15: Processo de avaliação de produto de software ISO/IEC14598-1

Fonte: (ABNT,1999)

2.1.4.4 ISO 13407

O processo de projeto centrado no usuário é uma filosofia de projeto que tem o usuário como foco principal, privilegiando desta forma a lógica de operação do sistema.

A norma ISO 13407 (1999) sobre “Processo de projeto centrado no usuário para sistemas interativos” fornece orientações sobre como atingir a qualidade em uso ao

incorporar atividades de projeto centrado no usuário em todo o ciclo de vida de sistemas interativos computacionais (Bevan, 2000).

Segundo Jokela (2003) o aspecto mais importante desta norma se refere à descrição das quatro principais atividades relacionadas ao projeto centrado no usuário (figura 16):

1) Compreender e especificar o contexto de uso: conhecer o usuário, o ambiente de uso e as tarefas que serão executadas com o produto.

2) Especificar os requisitos do usuário e os requisitos organizacionais: determinar os critérios de sucesso para a usabilidade do produto em termos das tarefas realizadas pelos usuários, determinar as diretrizes e limitações do projeto.

3) Produzir soluções de projeto: incorporar conhecimentos de IHC (design interativo, usabilidade) nas soluções de projeto.

4) Avaliar projetos em relação aos requisitos do usuário: a usabilidade do projeto deve ser avaliada em relação às tarefas dos usuários.

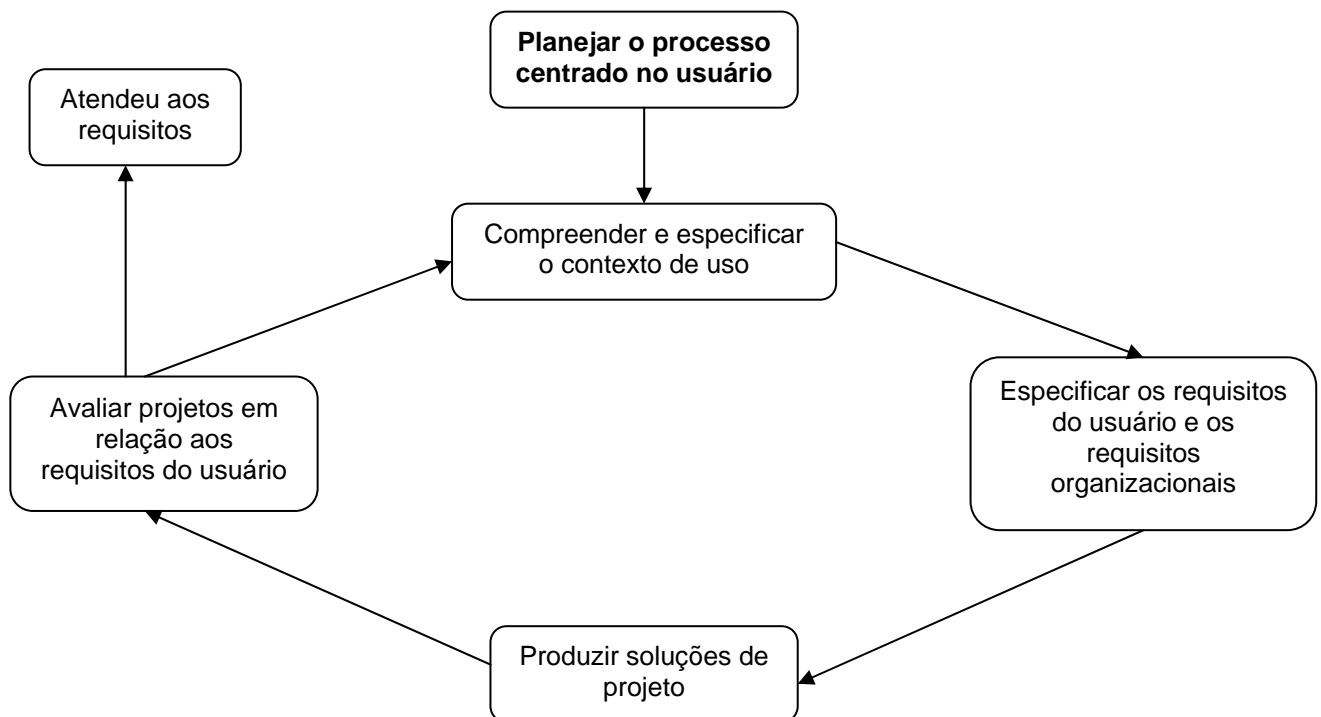


Figura 16: Ciclo de projeto centrado no usuário ISO 13407

Fonte: (Maguire, 2001)

A ISO 9241-11 e a ISO 13407 são duas normas muito importantes relacionadas à usabilidade (Jokela, 2003), a primeira fornece a definição de usabilidade enquanto a segunda fornece orientações sobre como projetar com usabilidade.

2.1.4.5 ISO/TR 18529

Finalmente, a ISO/TR 18529 (2000) contém um conjunto estruturado de processos, derivados da ISO 13407, referentes ao ciclo de vida do produto. Ela pode ser utilizada para avaliar a capacidade de uma organização de aplicar as atividades de Projeto Centrado no Usuário e também para especificar o que é necessário para implementar essa filosofia de processo (Bevan, 2000).

A figura 17 ilustra como as normas apresentadas anteriormente podem apoiar o processo de atingir a usabilidade de um produto. Estas normas podem ser categorizadas em relação a:

- Uso do produto (eficácia, eficiência e satisfação em relação a um determinado contexto de utilização).
 - Interface com o usuário e interação.
 - Processo utilizado para o desenvolvimento do produto.
 - Capacidade de uma organização de aplicar o processo de projeto centrado no usuário.

Ou seja, o objetivo é que o produto seja utilizado com eficácia, eficiência e satisfação quando usado em determinados contextos. Um pré-requisito para que isto ocorra é a existência de interface e interação adequadas. Isto exige que o processo de projeto seja centrado no usuário, o que, por sua vez, para ser atingido

consistentemente exige uma capacidade organizacional que apóie esta filosofia de projeto (Bevan, 2001).

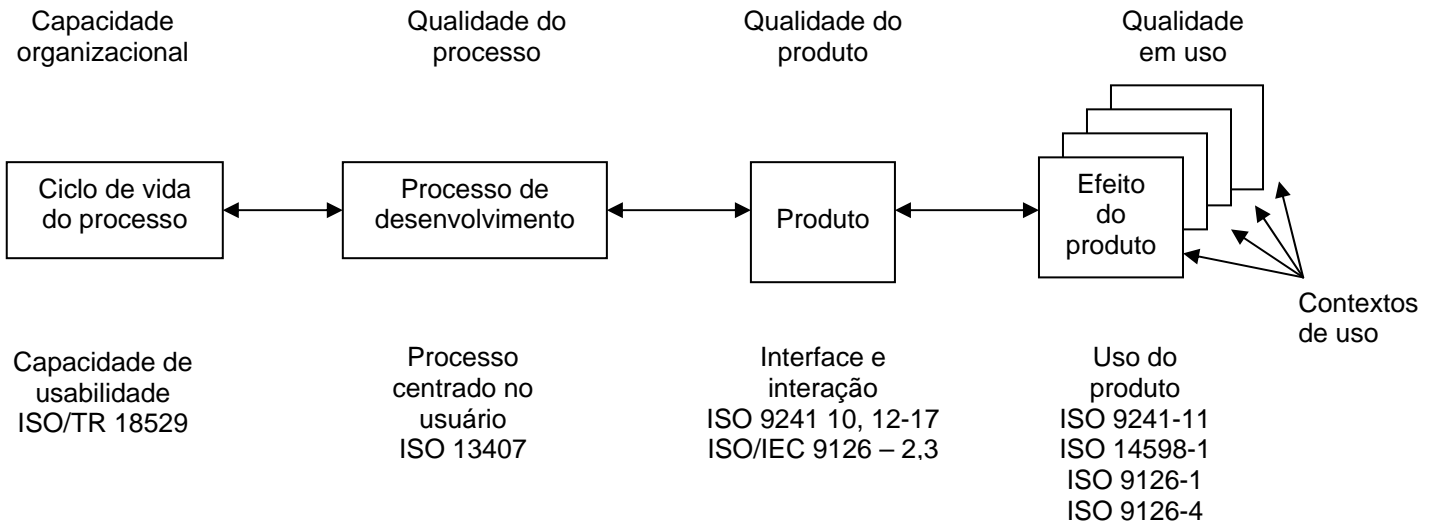


Figura 17: As normas e a usabilidade

Fonte: (Bevan, 2002)

2.1.5 Usabilidade segundo Hiltunen

Concluiremos as diversas abordagens sobre a definição de usabilidade com a visão de Hiltunen (2002) que aborda o tema sob a perspectiva da experiência do usuário móvel. Referindo-se ao serviço que está disponível ao usuário móvel através de PDA, telefone celular ou qualquer outro computador de mão, Hiltunen define o conceito de “experiência do usuário” em função de cinco componentes: utilidade, usabilidade, disponibilidade, estética e processo *offline* como ilustra a figura 18.

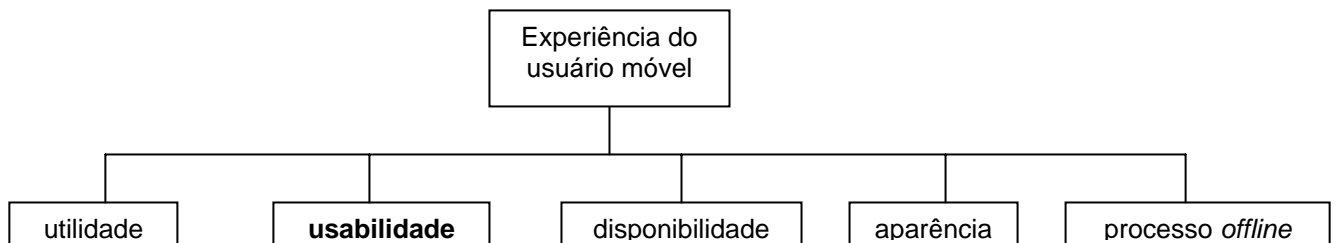


Figura 18: Componentes da experiência do usuário móvel

Fonte: (Hiltunen, 2002)

Hiltunen utilizou a definição da norma ISO 9241-11 para usabilidade e definiu utilidade como sendo a percepção do usuário em relação ao serviço lhe agregar algum valor. Além disso, o autor acrescentou três outros componentes que ele considera que exercem grande influência na opinião geral do usuário sobre o sistema: a disponibilidade, ou seja, o serviço deve estar sempre *online* e funcionando perfeitamente, a aparência deve ser agradável e atraente e todo o processo *offline* deve ser eficiente. Nesta última categoria o autor inclui vários elementos que estão por trás do serviço oferecido como o valor do nome da empresa que oferece o serviço, a confiança na empresa e no serviço oferecido e todo o processo de retaguarda como apoio ao usuário, velocidade de entrega de uma mercadoria, etc.

2.2 Definição de critérios para medir a usabilidade

Para Bevan (1999a) é imprescindível que a usabilidade passe a fazer parte da aceitabilidade de um produto como definido em (Nielsen, 1993a) e (Shackel, 1991). Entretanto, ao adquirir um produto tanto os usuários quanto as empresas têm pouca informação em relação a quão fácil é utilizar aquele produto, a quão produtivos eles serão ao utilizá-lo, ou a quanto treinamento e suporte serão necessários (Bevan, 1999b).

Em muitas situações a usabilidade é ignorada por falta da definição de critérios objetivos que permitam defini-la no desenvolvimento do produto (Bevan, 1999b). Segundo Good (1986 apud Öörni, 2003) se não houver especificações de usabilidade que possam ser mensuráveis não há nenhuma maneira de determinar os requisitos de usabilidade de um produto e tão pouco é possível medir se um

produto acabado atende a esses requisitos. Para o autor “se não é possível medir a usabilidade, não se tem a Engenharia de Usabilidade”.

Para Nielsen (1993a) medir a usabilidade é uma parte muito importante no ciclo da Engenharia de Usabilidade para verificar se os objetivos de usabilidade desejados foram efetivamente atingidos e também para comparar a usabilidade entre produtos concorrentes. Nielsen (1993a) define um modelo de medida da usabilidade ilustrado na figura 19.

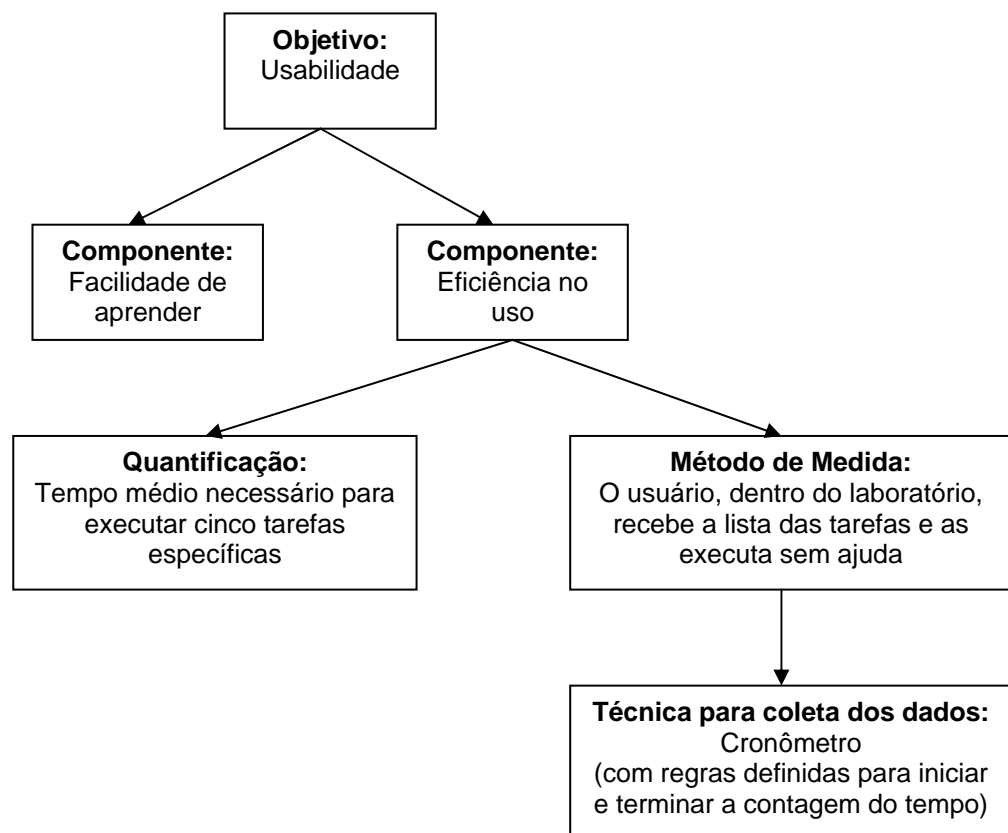


Figura 19: Modelo de medida da usabilidade segundo Nielsen

Fonte: (Nielsen, 1993a)

Neste modelo o primeiro passo é identificar claramente qual o objetivo a ser atingido, que neste caso é a usabilidade, mas poderia ser, por exemplo, aumentar a percepção dos clientes em relação à qualidade das interfaces dos sistemas

desenvolvidos por uma empresa. Os objetivos são normalmente conceitos abstratos que precisam ser definidos em termos de componentes, como por exemplo as componentes de usabilidade como definido em (2.1.2).

Uma vez definidas as componentes do objetivo, estas precisam ser quantificadas precisamente. O autor sugere uma série de medidas quantitativas que podem ser utilizadas como: tempo total para executar uma tarefa, o número de tarefas que podem ser executadas em um determinado intervalo de tempo, a quantidade de erros do usuário, a frequência com que o usuário utilizou alguma forma de ajuda e o tempo gasto com estes elementos do sistema, entre outras (Nielsen, 1993a).

A partir das medidas quantitativas é necessário definir um método para medir a performance do usuário. Uma vez que este método tenha sido definido é preciso determinar quais serão todas as atividades a serem executadas para coletar os dados.

2.2.1 O projeto MUSiC

Com o objetivo de definir um método e ferramentas para medir a usabilidade foi desenvolvido, a partir de uma iniciativa conjunta de oito empresas e entidades de pesquisa europeias, o projeto MUSiC – Measuring Usability of Systems in Context (Macleod, 1997).

Este projeto teve como principal objetivo o desenvolvimento de um método modular que pudesse ser utilizado para especificar os requisitos de usabilidade de um produto e para verificar se o produto atende a esses requisitos. O objetivo pretendido foi atingido pois o MUSiC é um método prático que apóia a Engenharia

de Usabilidade focando nas medidas de performance de usuários reais (Scapin, 1997).

São ao todo quatro tipos de métricas (analíticas, de performance, esforço cognitivo e satisfação do usuário) que, por serem complementares, permitem a medição dos diferentes componentes da usabilidade.

As métricas analíticas medem aspectos da usabilidade derivados das especificações do produto e dos modelos de usuários. Elas são utilizadas nas fases iniciais do desenvolvimento do projeto pois podem ser aplicadas em especificações que ainda estão no papel.

As métricas de performance medem com que eficácia e eficiência os usuários atingem seus objetivos ao utilizar o produto. Estas métricas são utilizadas nas várias versões de protótipos do produto e são baseadas nas análises da execução e do resultado da tarefa realizada pelo usuário.

As métricas de esforço cognitivo permitem que sejam obtidas medidas referentes ao esforço mental que foi exigido do usuário na utilização do produto. Estas métricas podem ser aplicadas durante todo o ciclo de projeto a partir da existência do primeiro protótipo.

Finalmente, é possível medir a satisfação subjetiva para avaliar a percepção do usuário sobre o produto. Estas métricas podem ser aplicadas a partir da versão beta do produto. Após utilizar o produto os usuários respondem a um questionário que vai fornecer uma visão sobre a sua atitude em relação ao sistema.

O MUSiC tem como referência a definição de usabilidade da ISO 9241-11. De acordo com esta definição a usabilidade, como sinônimo de qualidade em uso, pode ser medida na extensão em que um produto possa ser usado com eficácia, eficiência e satisfação dentro de um determinado contexto, ou seja, a usabilidade de um

produto é afetada não somente pelas características do produto, mas também pelo seu contexto de uso.

Na área de IHC há bastante tempo se reconhece que os usuários e as tarefas que eles executam têm grande efeito nos resultados de qualquer sistema de avaliação (Miller, 1971 apud Maguire, 2001). Na metade dos anos 80 houve um movimento maior em relação à importância do contexto com o trabalho de Whiteside (1988). O autor percebeu que embora alguns produtos apresentassem bons resultados em experimentos realizados dentro do laboratório, o mesmo não se repetia quando o produto era usado no mundo real. Estas pesquisas trouxeram novamente à tona a importância do contexto de uso na avaliação (Maguire, 2001).

O contexto de uso é definido a partir dos usuários do produto, das tarefas que eles têm que executar, dos ambientes técnico, organizacional e físico e até mesmo do dia e da hora em que o produto está sendo utilizado (Bevan, 2002). A figura 20 ilustra todos os componentes da usabilidade e as relações entre eles.

Desta forma, para especificar ou medir a usabilidade é necessário definir os objetivos de uso, o contexto de uso e as medidas de usabilidade (Bevan, 2002). Cada um destes itens está detalhado a seguir:

- **Objetivos:**

Primeiramente devem ser descritos os objetivos de uso do produto, ou seja, quais objetivos o usuário espera atingir ao utilizar aquele determinado produto. Estes objetivos podem ser decompostos em sub-objetivos e servirão, posteriormente, como orientação para a descrição das tarefas.

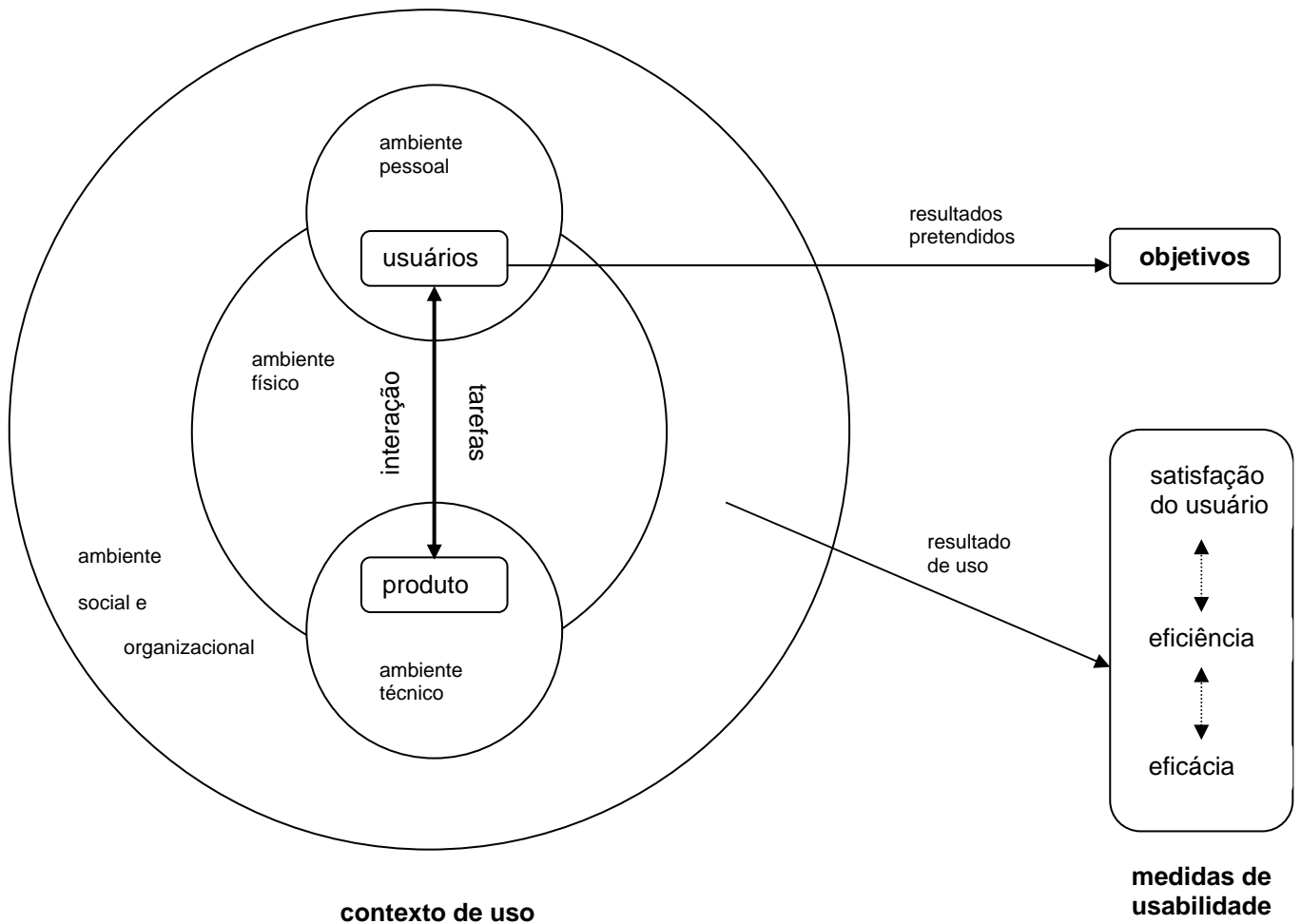


Figura 20: Componentes da usabilidade

Fonte: Adaptado de (Bevan, 2002) e (NBR 9241-11)

- Contexto de uso:

A análise do contexto de uso vai permitir a especificação das características dos usuários, das tarefas que eles executam e das circunstâncias reais de uso. A partir desta especificação é possível determinar quais atributos do contexto de uso irão afetar a usabilidade e portanto deverão ser analisados durante o processo de avaliação, assegurando desta forma a validação dos resultados (Thomas, 1996). O quadro 2 ilustra alguns dos principais atributos do contexto de uso.

Quadro 2: Atributos do contexto de uso

Sistema	Usuários	Tarefas
Sistema ou produto Nome Versão Descrição Finalidade Principais áreas de aplicação Principais funções Mercado alvo Para cada tipo de usuário Tipo Papel Objetivo Principais benefícios com o uso do sistema/produto Custos no uso do sistema/produto	Usuário Nome Tipo (primário/secundário/indireto) Experiência, habilidades, conhecimento Experiência com o produto Experiência relacionada ao uso do produto Conhecimento da tarefa Conhecimento organizacional Nível de treinamento Habilidade com os dispositivos de entrada de dados Qualificação Habilidades de linguagem Atributos pessoais Idade Sexo Capacidades físicas e cognitivas Limitações físicas e cognitivas Atitudes Motivação	Para cada tarefa da lista Nome Características Objetivos/resultados Passos Frequência Duração Flexibilidade Dependências Demanda física e mental Resultado Riscos resultantes de erros Demandas críticas de segurança
Ambiente		
Técnico	Físico	Organizacional
Hardware Software Rede Materiais de referência Outros equipamentos	Condições do local de trabalho Condições atmosféricas Ambiente acústico Ambiente térmico Ambiente visual Instabilidade ambiental Projeto do local de trabalho Espaço Móveis Postura do usuário Localização Segurança do local de trabalho Riscos para a saúde Roupas e equipamentos de proteção	Estrutura Grupos de trabalho Práticas de trabalho Assistência Interrupções Estrutura de gerência Estrutura de comunicação Salário Atitudes e cultura Política no uso dos computadores Objetivos organizacionais Relações industriais Projeto do trabalho Cargos Horas de trabalho Flexibilidade Monitoramento do desempenho Velocidade Autonomia Discrição

Fonte: (Maguire, 2001)

- Medidas de usabilidade:

A eficácia, a eficiência e a satisfação podem ser medidas de maneira objetiva a partir dos resultados da interação do usuário com o sistema. A importância de cada componente da usabilidade vai depender do contexto de uso e dos objetivos da avaliação, que vão determinar a escolha e o nível de detalhes de cada medida.

2.2.2 Etapas do Método de Medida da Performance

As métricas de performance do MUSiC estão descritas no “Método de Medida da Performance” (Cooper, 1995; Macleod, 1997; Thomas, 1996a; Thomas 1996b). Este método sugere um conjunto de medidas e métricas baseadas nas análises da execução e do resultado da tarefa realizada pelo usuário. As etapas necessárias para avaliar a usabilidade de um produto utilizando o Método de Medida da Performance (figura 21) estão descritas a seguir (Macleod, 1997; Bevan 1998).

1) Definir o produto a ser testado: identificar e descrever o produto, ou os componentes do produto que serão o foco da avaliação de usabilidade.

2) Definir o Contexto de Uso: identificar e descrever o contexto em que o produto será usado em função de: características dos atuais ou potenciais usuários, tarefas que serão executadas e ambiente em que os usuários utilizarão o produto.

3) Especificar os objetivos e o contexto de avaliação: definir quais requisitos de usabilidade serão analisado e garantir que a avaliação será conduzida em condições tão próximas quanto possível das condições reais de uso do produto. Para isso é necessário que os usuários sejam representativos da população alvo a qual é direcionado o produto, as tarefas sejam representativas em relação ao uso real do

produto e as condições da avaliação sejam representativas das condições normais nas quais o produto é utilizado.

4) Preparar a avaliação: preparar as tarefas, preparar os usuários, preparar o ambiente do teste.

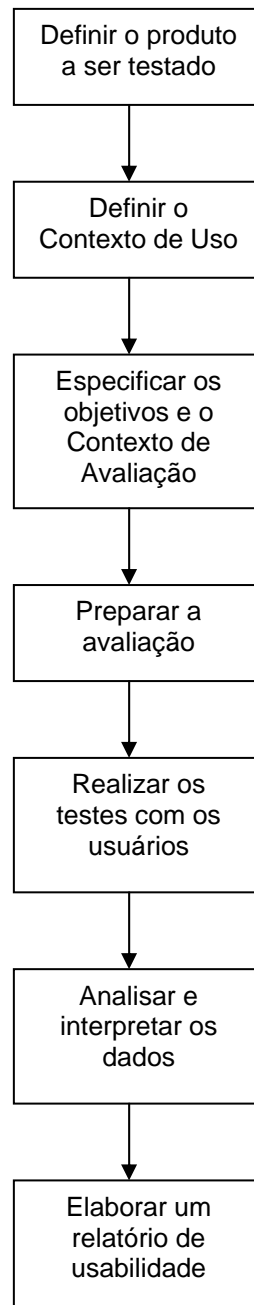


Figura 21: Etapas do Método de Medida da Performance

Fonte: (Coopers, 1995)

5) Realizar os testes com os usuários: os usuários serão monitorados e filmados enquanto executam determinadas tarefas. Ao avaliar a qualidade em uso é importante que os usuários trabalhem sozinhos, tendo acesso somente às formas de ajuda que estariam disponíveis se eles estivessem trabalhando em condições reais de uso.

6) Analisar e interpretar os dados: a partir das anotações feitas durante a sessão e da análise das fitas de vídeo é possível coletar as medidas e derivar as métricas de usabilidade.

7) Elaborar um relatório de usabilidade: o CIF-Common Industry Format for Usability Test Reports (CIF, 2001) foi desenvolvido com o objetivo de criar um padrão para a geração de relatórios de testes de usabilidade. Este guia traz todas as especificações técnicas que devem ser usadas ao se reportar a realização de um teste de usabilidade baseado na definição apresentada pela ISO 9241-11. Neste relatório estarão descritas as condições de realização da avaliação, os resultados obtidos e as recomendações sugeridas.

2.2.3 Medidas e métricas de usabilidade do MUSiC

A partir da análise do resultado da tarefa produzido pelo usuário e da fita de vídeo da sessão é possível obter medidas de performance a partir das quais derivarão as métricas de usabilidade (figura 22).

- **Medidas**

A análise do resultado da tarefa produzido pelo usuário permite medir a quantidade do resultado obtido e a qualidade do resultado obtido. A quantidade pode

ser medida como a proporção dos objetivos da tarefa que foram atingidos pelo usuário. A qualidade pode ser medida pela extensão com a qual o resultado da tarefa realizada pelo usuário corresponde aos critérios especificados.

A análise da fita de vídeo permite medir o tempo de duração da tarefa e o tempo improdutivo. O tempo de duração da tarefa corresponde ao tempo total desde o início da interação até o usuário considerar a tarefa como concluída. O tempo improdutivo corresponde a todo tempo que o usuário gastou executando ações que não contribuíram para o resultado da tarefa.

- **Métricas**

A partir das medidas definidas anteriormente é possível derivar as seguintes métricas:

- Eficácia da tarefa: a qualidade e a quantidade com que os usuários atingem os objetivos específicos.
- Eficiência do Usuário: relaciona a eficácia ao tempo gasto para executar a tarefa.
- Eficiência Relativa: relaciona a eficiência do usuário comparada a eficiência de um usuário especialista executando a mesma tarefa nas mesmas condições.
- Período Produtivo: a proporção de tempo gasto pelo usuário executando ações que contribuíram para o resultado da tarefa.

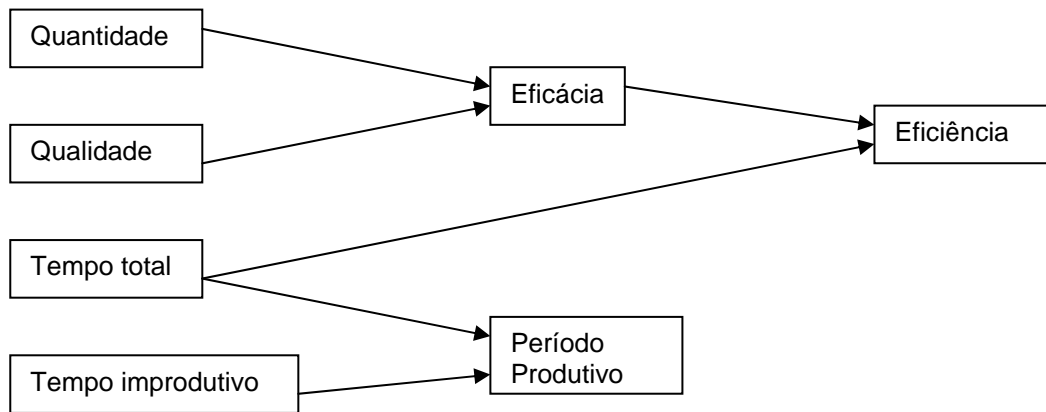


Figura 22: Relações entre as medidas e métricas de usabilidade no MUSiC

Fonte: (Coopers, 1995)

2.2.3.1 Eficácia

A eficácia é uma medida que define o quão bem os usuários atingiram os objetivos da tarefa, independentemente do tempo gasto e das dificuldades encontradas durante a realização da tarefa. Ela é definida em função de duas componentes, a qualidade e a quantidade, e é determinada pela fórmula:

$$\text{Eficácia} = (\text{quantidade} \times \text{qualidade}) / 100\%$$

A quantidade corresponde ao percentual de objetivos atingidos pelo usuário. Para medi-la é preciso: definir como será representado o resultado da tarefa pelo usuário, definir o que corresponde a um objetivo atingido, contar o número de objetivos atingidos pelo usuário e verificar o percentual desse valor em relação ao número total de objetivos.

A qualidade corresponde à extensão com a qual o resultado da tarefa realizada pelo usuário corresponde aos critérios especificados. Para medi-la é preciso: definir o que corresponde a um objetivo ideal a ser atingido, definir um sistema de

pontuação que atribua valores aos objetivos atingidos pelo usuário quando comparados aos objetivos ideais e calcular a qualidade total como o percentual da soma desses pontos em relação à pontuação máxima possível.

2.2.3.2 Eficiência

2.2.3.2.1 Eficiência temporal

A eficiência pode ser medida relacionando os níveis de eficácia atingidos em relação ao tempo gasto na execução da tarefa. Esta métrica é definida pela fórmula:

$$\text{Eficiência} = \text{eficácia} / \text{tempo da tarefa}$$

2.2.3.2.2 Eficiência relativa

A eficiência relativa do usuário é definida como a razão entre a eficiência de qualquer usuário e a eficiência de um usuário especialista e é determinada pela fórmula:

$$\text{Eficiência relativa} = (\text{eficiência do usuário} / \text{eficiência do especialista}) \times 100\%$$

O usuário especialista é definido como qualquer pessoa que tenha sido treinada no uso do sistema e que tenha uma alta experiência no seu uso.

A eficácia e o tempo de execução da tarefa do usuário especialista devem ser medidos enquanto ele executa as mesmas tarefas, com o mesmo produto e nas mesmas condições que o usuário que está participando da avaliação.

Se houver mais de um usuário especialista disponível, a eficiência pode ser considerada como a média entre a eficiência de cada usuário especialista. Quanto mais especialistas forem considerados, mais fiel será essa média.

Esta métrica pode indicar a posição do usuário na curva de aprendizagem do sistema.

2.2.3.3 Período produtivo

O período produtivo não é medido diretamente, ele é calculado subtraindo do tempo total o tempo gasto em ações improdutivas. O método classifica essas ações em três categorias:

Tempo de Ajuda: tempo gasto pelo usuário na obtenção de informações sobre o sistema.

Tempo de Busca: tempo gasto pelo usuário explorando a estrutura do sistema sem ativar nenhum componente.

Tempo de *Snag*: são as ações que não contribuem direta ou indiretamente com o resultado da tarefa, mas que não podem ser classificadas como ações de ajuda ou busca.

O período produtivo é então definido por:

$$\text{Período produtivo} = \frac{(\text{tempo total} - \text{tempo ajuda} - \text{tempo busca} - \text{tempo snag})}{\text{tempo total}} \times 100\%$$

Pode ocorrer que a eficácia seja zero, quando o usuário não conseguir atingir nenhum percentual dos objetivos, ou quando a qualidade dos objetivos atingidos não corresponder ao que foi especificado. Neste caso, há uma regra no Método de Medida da Performance que estabelece que o período produtivo também deve ser

considerado zero, mesmo que tenham ocorrido períodos produtivos durante a execução da tarefa. Esta regra foi incluída para não haver contradição com a definição para a métrica de período produtivo, segundo a qual o período produtivo corresponde à proporção de tempo em que o usuário esteve progredindo em direção aos objetivos da tarefa. Como os objetivos não foram atingidos, não houve progresso algum e conseqüentemente não pode ser considerado nenhum período produtivo.

2.2.3.4 Medidas de satisfação

A performance e a satisfação do usuário não estão diretamente relacionadas, o usuário pode apresentar uma ótima performance embora não goste do sistema e vice-versa (Macleod, 1997).

Muitos aspectos da usabilidade podem ser melhor estudados simplesmente fazendo perguntas ao usuário, principalmente quando estes aspectos estão ligados a questões relacionadas à satisfação subjetiva do usuário. As medidas de satisfação descrevem a usabilidade como ela é percebida pelo usuário em relação ao sistema como um todo.

O uso de questionários e entrevistas é um método útil para avaliar a opinião do usuário. Este método também pode ser utilizado para avaliar possíveis ansiedades dos usuários pois, mesmo que seja possível fazer uso de medidas objetivas como batimentos cardíacos e freqüência respiratória, ou até mesmo da observação de postura e movimentos do corpo do usuário, estas medidas dificilmente são utilizadas nos projetos de desenvolvimento de *software* (Nielsen, 1993a).

Os questionários podem conter questões abertas, mais raras de serem respondidas pelo usuário, ou questões fechadas (Nielsen, 1993a). Normalmente eles são compostos por escalas que podem ser escalas Likert, ou escalas de diferencial semântico (Jordan, 1998).

As escalas Likert apresentam uma frase à qual o usuário deve indicar o seu grau de concordância em uma escala de 1-5 ou 1-7 (figura 23).

	1-Discordo completamente	2-Discordo parcialmente	3-Neutro	4-Concordo parcialmente	5-Concordo completamente
É muito agradável usar este sistema.					

Figura 23: Exemplo de escala Likert

A escala de diferencial semântico apresenta dois termos opostos ao longo de uma dimensão e pede que o usuário selecione a posição na escala que melhor reflete a sua opinião (figura 24).

Eu considero este sistema:	
Agradável	----- Irritante
Simple	----- Complicado
Rápido	----- Lento

Figura 24: Exemplo de escala de diferencial semântico

Alguns questionários que podem ser aplicados em uma avaliação de usabilidade são o SUMI-Software Usability Measurement Inventory (SUMI, 2003), o QUIS-Questionnaire for User Interaction Satisfaction (QUIS, 2003) e o SUS-System Usability Scale (SUS, 2003; Brooke, 1996).

2.3 As técnicas de usabilidade no ciclo de desenvolvimento do produto

Diferentes técnicas de usabilidade podem ser empregadas em função das diferentes fases do ciclo de desenvolvimento do produto. Os resultados das aplicações dessas técnicas fornecem subsídios importantes para a tomada de decisão sobre necessidades de alteração de projeto a fim de aumentar a usabilidade. É sempre recomendado que sejam feitas avaliações menores em várias versões de protótipos em lugar de realizar uma única avaliação já com versões finais do produto (Nielsen, 1993a).

Quanto mais cedo forem realizadas as avaliações de usabilidade, mais simples será a alteração no projeto e conseqüentemente mais facilmente os requisitos de usabilidade podem ser incorporados ao produto. Esta abordagem iterativa (figura 25) reduz custos e evita reformulações muito grandes ou às vezes até impossíveis de serem realizadas quando o produto já se encontra em suas versões finais.

São diversas as técnicas de usabilidade que podem ser utilizadas durante o desenvolvimento de um projeto, o trabalho de Ivory (2001) lista 39 técnicas diferentes. Não existe uma única técnica capaz de identificar todos os problemas de usabilidade de uma interface. As técnicas são complementares e a sua escolha dependerá principalmente das diferentes fases do ciclo de desenvolvimento do projeto, dos objetivos da avaliação e dos recursos humanos, físicos e financeiros que estiverem disponíveis.

As técnicas diferem quanto ao tipo e quantidade de problemas que identificam, à sistematização de seus resultados, à facilidade de aplicação e às chances que seus resultados apresentam para convencer os projetistas das necessidades de mudanças na interface (Cybis, 2002).

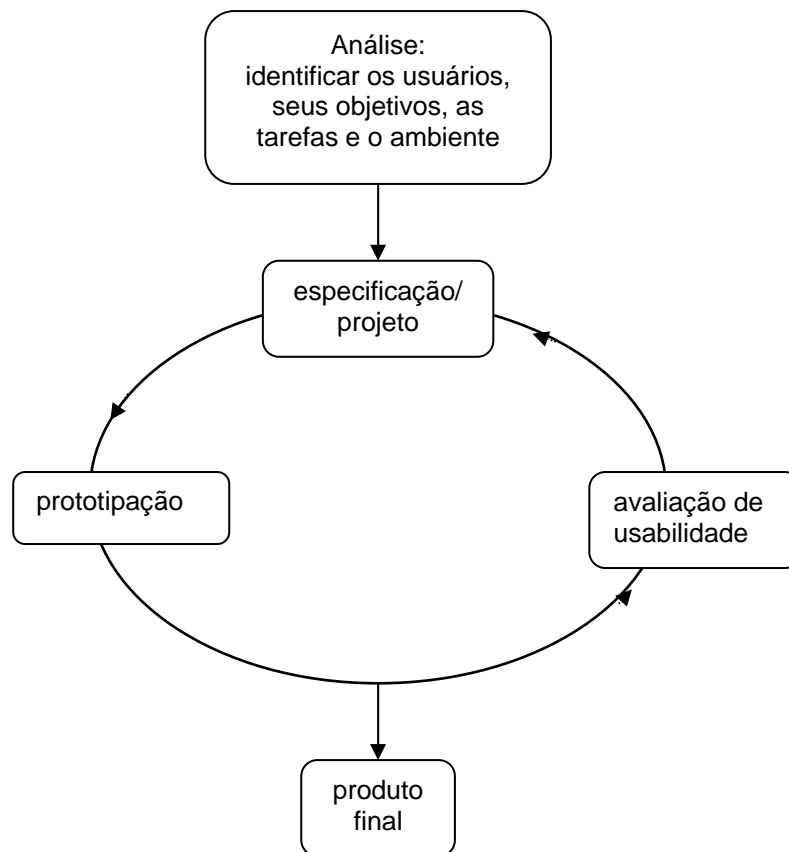


Figura 25: Avaliação de usabilidade no desenvolvimento de um produto

Fonte: Adaptado de (Weiss, 2002)

Os resultados de uma avaliação de usabilidade fornecem subsídios para novas especificações de projeto, daí a importância das avaliações estarem presentes desde o início do projeto.

Na fase inicial de análise de requisitos técnicas como *focus group* (Jordan, 1998), *Contextual Inquiry* (Hackos, 1998), questionários e entrevistas podem ser utilizadas. Para envolver o usuário durante o desenvolvimento do projeto podem ser utilizadas técnicas como design participativo (Hackos, 1998), sessões de arranjo e classificação ou diagramas de afinidade (Gaffney, 2000).

Cybis (2002) classifica as técnicas de avaliação de usabilidade em três categorias:

- **Prospectivas:** buscam a opinião do usuário sobre a interação com o sistema. São baseadas em questionários e entrevistas com os usuários para avaliar a sua satisfação ou insatisfação com o sistema (Jordan, 1998).
- **Preditivas/Analíticas:** buscam prever os erros de projeto de interfaces sem a participação direta de usuários. As avaliações são baseadas em verificações e inspeções feitas por especialistas em usabilidade ou projetistas. São exemplos de técnicas preditivas (Nielsen, 1994): Análise Hierárquica da Tarefa, Avaliação Heurística, Inspeções Ergonômicas via Checklists, Inspeção Cognitiva.
- **Objetivas/Empíricas:** buscam constatar os problemas a partir da observação do usuário interagindo com o sistema. São os Ensaios de Interação, também chamados Testes de Usabilidade, que serão descritos em mais detalhes a seguir.

2.3.1 Ensaios de interação

Os ensaios de interação são uma simulação de uso do sistema. Usuários representativos executam tarefas determinadas com uma versão do produto, que pode variar desde um protótipo em papel até a sua versão final.

Segundo Nielsen (1993a, p.165) “realizar ensaios de interação com usuários reais é o método de usabilidade mais fundamental e, de certa forma, insubstituível uma vez que proporciona informações diretas sobre como as pessoas estão usando os computadores e quais problemas elas estão encontrando com as interfaces concretas que estão sendo testadas”.

Macleod (1997, p.292) recomenda que sejam feitas observações do usuário sempre que possível pois “nenhum processo de projeto centrado no usuário pode estar completo sem a realização de algum tipo de teste com os usuários”.

Enquanto realizam as tarefas, os usuários são observados por avaliadores que tomam notas sobre as ações e reações dos usuários. O ensaio pode ser gravado utilizando câmeras de áudio e vídeo, gerando fitas que serão analisadas posteriormente.

Um aspecto importante a ser considerado é quanto ao local de realização do ensaio de interação que pode ser um laboratório de usabilidade ou o próprio local em que o usuário realiza a tarefa.

O laboratório é um local próprio para a realização das avaliações, equipado com várias câmeras que podem estar focalizando o usuário e os demais componentes do equipamento (figura 26). A maioria destes laboratórios tem uma sala especial à prova de som, equipada com um espelho falso que permite que o participante realize o teste sozinho enquanto o avaliador o observa através do vidro. O uso de várias salas permite que outras pessoas envolvidas no projeto também assistam à avaliação, sem perturbar o participante. É importante mostrar a sala de observação para o participante, pois é bem menos estressante saber quem e o que se esconde atrás do espelho que ficar imaginando o que está se passando. Normalmente o participante acaba ignorando a presença do observador durante o teste, mesmo sabendo que ele está ali (Nielsen, 1993a).

Quando o teste é feito no próprio local de realização da tarefa é possível observar o usuário interagindo com o sistema em uma situação real de uso, ou seja, com todos os ruídos e interferências alheias à tarefa que exercem influência na interação. Estas são informações valiosas que dificilmente apareceriam em um teste realizado dentro do laboratório. O avaliador pode realizar a observação utilizando somente lápis e papel para fazer as anotações, deixando o usuário à vontade e procurando interferir o mínimo possível em sua rotina.

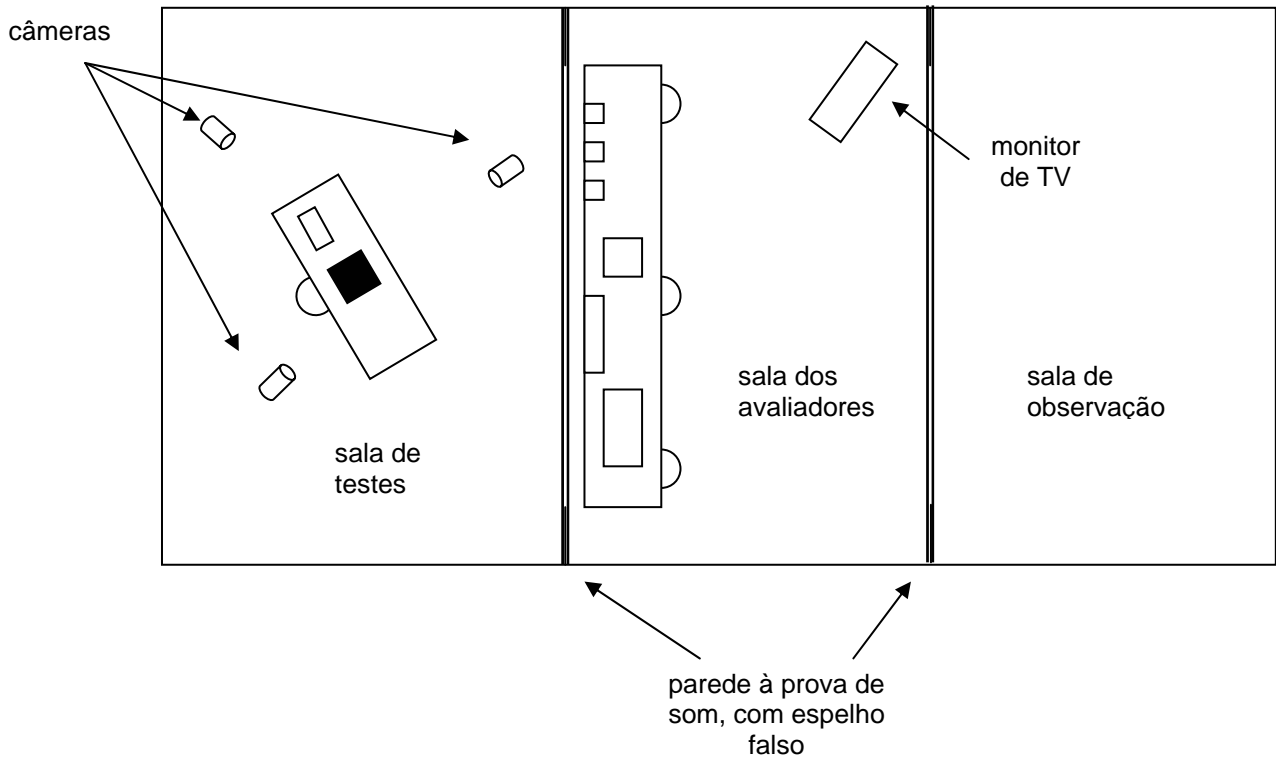


Figura 26: Modelo de *layout* de um laboratório de usabilidade

Fonte: Adaptado de (Nielsen, 1993a)

Para obter informações mais detalhadas é possível também utilizar laboratórios portáteis compostos normalmente de *laptop*, câmera de vídeo, microfone e *software* para análise, que permitem que o local de trabalho do usuário rapidamente seja convertido em um local de teste (USI, 2002).

Para determinar o que os usuários estão pensando os analistas podem utilizar a técnica de verbalização simultânea (*think aloud*), na qual é solicitado ao usuário que comente em voz alta o que ele está pensando enquanto interage com a interface. O analista pode fazer algumas perguntas durante a interação para esclarecer algumas atitudes do usuário, desta forma é possível saber não só o que o usuário está fazendo, mas o porquê de sua atitude. No entanto alguns autores consideram que esta técnica pode interferir na performance do usuário, pois ao ter que verbalizar suas atitudes e pensamentos sua atenção é desviada da tarefa (Cybis, 2002;

Nielsen, 1993a; Jordan, 1998). Uma alternativa é a realização da verbalização consecutiva (*debriefing*) que é uma entrevista com o usuário realizada imediatamente após o término do ensaio, na qual ele comenta sobre as tarefas que acabou de executar. Para que o usuário se lembre mais facilmente de suas atitudes podem ser utilizadas as fitas de vídeo como forma de apoio.

Os analistas não devem interferir nas ações dos usuários e devem tomar todas as medidas necessárias para que este se sinta à vontade durante a realização dos ensaios de interação. Para que isto ocorra é importante esclarecer ao usuário os objetivos do teste, as tarefas que ele irá realizar, como o teste será conduzido e qual será sua duração.

É de extrema importância deixar bem claro ao usuário que o que está sendo avaliado é a interface e não ele. Ao participar de um ensaio de interação o usuário pode sentir como se ele estivesse sendo testado o que pode deixá-lo constrangido e com medo de cometer erros. Essa atitude, que pode vir a atrapalhar a avaliação, pode ser evitada se o papel do usuário for colocado de forma bem clara no início da sessão: ele está ali para auxiliar na avaliação da interface e não para ser testado.

Os resultados de um ensaio de interação podem ser muito úteis para convencer as pessoas envolvidas no projeto que a interface apresenta problemas. É mais fácil convencer projetistas e desenvolvedores a conduzir reformulações no projeto quando eles têm a oportunidade de observar pessoalmente os usuários enfrentando dificuldades no uso da interface (Nielsen, 1993a; Macleod, 1997).

Esta é uma das técnicas mais recomendadas para avaliar a usabilidade, pois segundo Jordan (1998), nada pode substituir a observação dos usuários tentando usar um produto.

2.3.1.1 Número de usuários em um ensaio de interação

O número de participantes necessários para realizar um ensaio de interação é um assunto ainda muito polêmico. Uma das primeiras pesquisas sobre este assunto foi o trabalho de Virzi (1992) que modelou o número de problemas de usabilidade encontrados em função do número de usuários utilizados na avaliação. Nesta pesquisa o autor chegou a três conclusões: (a) 80% dos problemas de usabilidade são encontrados por quatro ou cinco participantes, (b) à medida que aumenta o número de participantes estes revelam cada vez menos informações novas, (c) os problemas de usabilidade mais graves têm maior probabilidade de serem revelados pelos primeiros poucos participantes.

Uma pesquisa semelhante foi realizada por Nielsen (1993c) que constatou a partir de vários estudos realizados, que em média 31% dos problemas de usabilidade são encontrados por um único usuário. Com base nesse valor o autor conclui que cinco usuários são suficientes para encontrar 85% dos problemas de usabilidade e que com quinze usuários podem ser encontrados 100% dos problemas. O autor considera perda de tempo observar os mesmos problemas se repetirem a partir do quinto usuário e portanto recomenda que se houver a disponibilidade de recursos para testar com 15 usuários é melhor distribuir esses recursos em três testes com cinco usuários. Nielsen (2000) argumenta que o principal objetivo da avaliação é melhorar a interface e não simplesmente documentar os seus pontos fracos.

Apesar dessas duas pesquisas serem freqüentemente citadas para justificar um pequeno número de usuários nos ensaios de interação seus resultados são questionados por vários autores (Caulton, 1999; Spool, 2001; Woolrych, 2001).

Em seu estudo Spool (2001) encontrou somente 35% dos problemas de usabilidade com os primeiros cinco usuários. Entretanto, como o próprio autor reconhece, este estudo foi realizado em um ambiente mais complexo em relação aos estudos de Virzi e Nielsen, pois os usuários navegaram em sites Web de *e-commerce* com a tarefa de comprar qualquer coisa que desejassem desde que não ultrapassassem um determinado valor em dinheiro estipulado previamente. As particularidades de cada participante influenciaram claramente estes resultados pois nenhum usuário comprou o mesmo produto que outro e nenhum percorreu os sites da mesma maneira. Segundo o autor, para avaliar a complexidade das interfaces Web cinco usuários não é um número suficiente.

Para Caulton(1999) e Woolrych(2001) o número cinco só pode ser aplicado quando a probabilidade de um usuário encontrar um problema for alta (31% ou mais) e similar para todos os usuários, ou seja, as diferenças individuais entre o perfil dos usuários deve ser mínima, pois quanto mais heterogêneos forem os grupos, mais usuários serão necessários para que todos os subgrupos possam estar representados. Para Caulton (1999) os resultados de Nielsen (1993c) podem ser aplicados a testes de usabilidade mais simples que têm como objetivo simplesmente detectar a presença, e não a prevalência estatística, de problemas de usabilidade.

Se os testes forem realizados com poucos participantes muitos problemas sérios da interface podem passar despercebidos, entretanto testar com muitos usuários pode ser inviável pois cada participante a mais representa custos extras em termos de tempo na execução das sessões e na posterior análise dos dados, e de recursos financeiros para seleção e remuneração do participante e do avaliador (Woolrych, 2001).

Uma relação custo/benefício deve ser estabelecida para determinar quantos usuários deverão testar um sistema em função principalmente da complexidade da interface, dos tipos e da quantidade de seus problemas, da população alvo, da disponibilidade de recursos e do tempo disponível.

Entretanto todos os autores são unânimes em afirmar que é melhor realizar alguns testes, ainda que com poucos participantes, a não realizar nenhum tipo de avaliação com usuários (Nielsen, 2000; Spool, 2001; Perfetti, 2003).

2.3.1.2 Validade e confiabilidade dos ensaios de interação

Um aspecto muito importante que deve ser considerado na realização de uma avaliação de usabilidade diz respeito à confiabilidade e à validade da técnica utilizada (Nielsen, 1993a).

A confiabilidade se refere ao grau de certeza de que o resultado obtido na avaliação será o mesmo se o teste for repetido (Nielsen, 1993a). Utilizar um maior número de usuários e realizar as análises dos dados empregando técnicas estatísticas podem aumentar a confiabilidade de uma avaliação (Hughes, 1999; Nielsen, 1993a).

A validade se refere a verificar se o teste de usabilidade está realmente avaliando aspectos relevantes da usabilidade do produto que se deseja testar (Nielsen, 1993a). Deseja-se assegurar que o resultado obtido seja realmente significativo considerando a utilização do produto real fora do laboratório em condições normais de uso (Rocha, 2003). A correta especificação do contexto de uso guiando a escolha dos usuários, das tarefas e do ambiente de uso podem assegurar a validade de uma avaliação.

2.4 Critérios para comparação entre técnicas de avaliação de usabilidade

Muitos trabalhos já foram realizados com o objetivo de comparar diferentes técnicas de avaliação de usabilidade para os computadores de mesa, verificando vantagens e desvantagens entre cada uma delas (Bastien, 1995; Desurvire, 1994; Karat, 1994; Sutcliffe, 1997; Sears, 1997; Jeffries, 1991, Savage, 1996). Serão apresentados aqui três destes trabalhos que resultaram na definição de um conjunto de medidas e métricas que podem ser utilizadas quando se deseja comparar diferentes técnicas de avaliação de usabilidade.

Bastien(1995) definiu três características que devem ser consideradas ao se comparar técnicas de avaliação de usabilidade: validade, completude e confiabilidade:

- Validade: é definida pela habilidade do avaliador em focar aspectos específicos e pré-definidos do design.
- Completude: a avaliação é considerada completa se ela consegue examinar o máximo possível da interface.
- Confiabilidade: a avaliação é considerada confiável se for possível reproduzir os mesmos resultados com outra avaliação realizada nas mesmas condições.

Como o autor não indicou como essas características poderiam ser medidas, Sears (1997) propôs estender a definição desses conceitos e definiu um conjunto de métricas que permitem compará-los quantitativamente (cada métrica tem uma escala de 0 a 1, onde 1 é o valor ótimo):

- Validade: uma técnica é considerada válida se permite que o avaliador foque nos aspectos que são pertinentes, ou seja, o avaliador não vai identificar como um problema de usabilidade um aspecto que não cause nenhum impacto no usuário. Desta forma a validade pode ser definida como a razão entre a quantidade de problemas reais identificados e a quantidade de todos os problemas identificados como problemas de usabilidade.

$$\text{Validade} = \frac{\text{quantidade de problemas reais encontrados}}{\text{quantidade de problemas identificados como problemas de usabilidade}}$$

A validade mede quanto esforço extra está sendo gasto em aspectos que não são importantes, ou seja, na identificação de problemas que não são considerados problemas reais de usabilidade.

- Completude: a técnica é considerada completa se permite que os avaliadores avaliem todos os aspectos da interface. Desta forma a completude pode ser definida como a razão entre os problemas reais que são identificados e os problemas que existem no sistema.

$$\text{Completude} = \frac{\text{quantidade de problemas reais encontrados}}{\text{quantidade de problemas reais que existem}}$$

A completude mede o percentual de problemas encontrados.

- Confiabilidade: a confiabilidade implica em que resultados similares devem ser obtidos quando as condições da avaliação forem reproduzidas. Desta forma a confiabilidade pode ser definida como a razão entre o desvio padrão da quantidade

de problemas encontrados e a média da quantidade de problemas encontrados nas várias avaliações.

$$\text{Confiabilidade} = 1 - \frac{\text{desvio padrão(quantidade de problemas reais)}}{\text{média(quantidade de problemas reais)}}$$

Para eliminar a possibilidade de um valor negativo para a confiabilidade, o que ocorre sempre que o numerador for maior que o denominador, a confiabilidade será definida como:

$$\text{Confiabilidade} = \text{Valor Máximo}(0, \text{Confiabilidade})$$

A confiabilidade mede a consistência entre as diferentes avaliações.

O trabalho de Jeffries (1991) também apresenta um conjunto de medidas que podem ser utilizadas na comparação entre as diferentes técnicas de avaliação. As medidas derivam de um relatório que deve ser feito pelo avaliador com todos os problemas de usabilidade que ele venha a encontrar durante a avaliação. Para o autor um problema de usabilidade é definido como “qualquer coisa que impacte a facilidade de uso, desde um problema interno do sistema até uma palavra grafada de maneira incorreta” (Jeffries, 1991, p.121).

A partir do relatório com todos os problemas de usabilidade encontrados pelo avaliador, o autor sugere que sejam definidas as seguintes medidas:

- Quantidade de problemas identificados pelo avaliador.
- Quantidade de problemas não duplicados, ou seja, problemas que só foram identificados naquela avaliação específica.
- Severidade dos problemas encontrados variando de 1 (problema trivial) a 9 (problema crítico). O grau de severidade deve ser atribuído pelo avaliador em função

do impacto do problema, da frequência com que ele ocorreu e do número de usuários que seriam afetados por ele.

- Número de horas gastas na avaliação.
- Benefício/custo: com base nas duas medidas anteriores, grau de severidade e horas gastas na avaliação (tempo gasto na aplicação da técnica), o autor define uma métrica para avaliar a razão entre o benefício e o custo:

$$\text{Benefício/custo} = \frac{\text{soma dos graus de severidade dos problemas encontrados}}{\text{pessoa-hora de avaliação}}$$

O autor também sugere que os problemas identificados sejam classificados em uma das três seguintes categorias:

- Consistência: o problema revela um aspecto da interface que está em conflito com outras partes do sistema?
- Recorrência: o problema interfere na interação somente na primeira vez em que ele ocorre ou ele será sempre um problema?
- Generalidade: o problema aponta uma falha geral que afeta várias partes da interface ou é um problema específico de uma única parte?

Com esses parâmetros Jeffries (1991) sugere que pode ser feita uma análise comparativa entre as técnicas em função de qual técnica apontou o maior número de problemas, qual teve maior ou menor razão benefício/custo, qual apontou os problemas de maior grau de severidade e qual apontou mais problemas gerais, recorrentes e de consistência.

3 AVALIAÇÃO DE USABILIDADE PARA OS COMPUTADORES DE MÃO

Os estudos apresentados neste capítulo ilustram como estão sendo conduzidas as avaliações de usabilidade para os computadores de mão. A partir destas pesquisas é possível analisar quais as principais técnicas que estão sendo utilizadas, como estão sendo definidos os diferentes contextos de avaliação, e como os pesquisadores vêm medindo a usabilidade nestes contextos.

3.1. Avaliação de usabilidade com emuladores

Quando se avalia a usabilidade de sites Web é recomendado que sejam feitos os testes em várias versões e modelos de *browsers*. Da mesma forma, para avaliar as aplicações desenvolvidas para os computadores de mão, seria necessário ter acesso a todos os equipamentos nos quais ela poderia ser utilizada, pois o conteúdo se apresenta de forma diferente em cada um deles. Entretanto, há uma enorme variedade de marcas e modelos de computadores de mão disponível no mercado, sem considerar que novos modelos de equipamentos, com novas características e mais recursos, continuam surgindo a todo momento. Como é impossível que o desenvolvedor tenha acesso a todos os tipos de equipamentos, fabricantes e desenvolvedores de *browsers* disponibilizam emuladores que possibilitam testar a aplicação no computador (figura 27). O emulador pretende simular o funcionamento do equipamento real e em alguns casos também a sua aparência, que pode ser

simulada quando for possível trocar a imagem padrão do equipamento que o emulador apresenta pela imagem do equipamento real que se deseja avaliar.

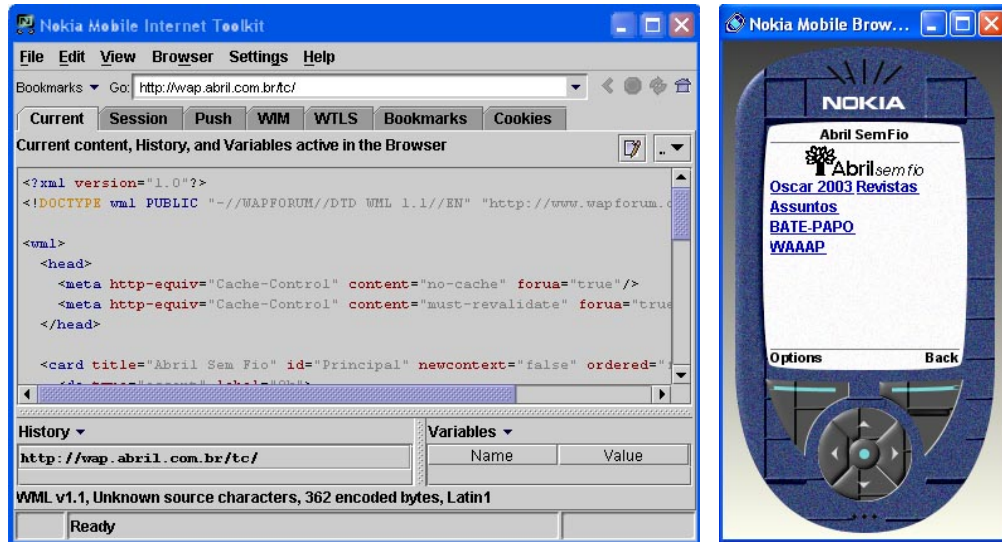


Figura 27: Exemplo de emulador no computador
(Emulador Nokia Mobile Internet Toolkit v3.1)

Fonte: (Nokia, 2003) © Copyright 2003 Nokia. Todos os direitos reservados.

Kaasinen (2000) realizou um estudo sobre como adaptar serviços disponíveis na Internet para que possam ser acessados por computadores de mão. O autor utilizou emuladores no computador para realizar duas avaliações de usabilidade de dois serviços WAP diferentes. A avaliação da primeira aplicação envolveu seis usuários, entre 13 e 52 anos, com experiências variadas quanto ao uso de telefone celular e ao uso de Internet. Em sessões com duração de duas horas o usuário avaliou a aplicação utilizando dois emuladores diferentes: um emulador de telefone celular que o usuário manipulava diretamente pela tela do computador e um emulador de PDA operado com o mouse. A avaliação da segunda aplicação foi realizada com quatro usuários, entre 22 e 25 anos, todos com experiência no uso de telefone celular e Internet, as sessões tiveram duração de uma hora e foram realizadas com um emulador de telefone celular operado com o mouse. Essas avaliações tiveram

como objetivos: saber a opinião dos usuários sobre o sistema, a utilidade e a facilidade de uso da aplicação, e identificar como era realizada a navegação entre as páginas. Os dados resultantes das avaliações foram baseados nas anotações dos avaliadores e nos diversos comentários e sugestões dos participantes. A partir dos resultados obtidos o autor sugere algumas *guidelines* para a conversão de aplicações de Internet para equipamentos WAP. Para Kaasinen, o emulador foi bastante útil pois permitiu obter a opinião geral dos usuários em relação às aplicações WAP o que resultou em alterações no *software* visando a atingir os requisitos apontados pelos participantes.

Buchanan (2001) fez uso do emulador em seu estudo com dois propósitos diferentes: avaliar qual a opinião geral dos usuários sobre os serviços WAP e avaliar a usabilidade de diferentes métodos para apresentar texto na tela do telefone celular. Para avaliar qual a percepção do usuário em relação aos serviços WAP foram selecionados 110 estudantes universitários, todos com experiência no uso de telefone celular, mas nenhum com experiência no uso de serviços ou equipamentos WAP. Usando um emulador no computador cada participante deveria interagir com três sites WAP, navegando em busca de informações que eles achassem interessantes. Após estarem familiarizados com os três sites, os participantes respondiam a um questionário de satisfação que deveria ser preenchido no próprio computador e enviado via Internet. Os resultados mostraram que os usuários perceberam o modelo WAP como algo simples e fácil de ser compreendido, entretanto não estavam dispostos a utilizar este serviço se a navegação fosse complexa, monótona e propensa a erros. O emulador foi então novamente utilizado para avaliar qual o melhor método para apresentar texto na tela entre três alternativas: rolagem horizontal (ao colocar o cursor sobre a linha, aparecem os

caracteres que antes não eram visíveis) , rolagem vertical (todos os caracteres são visíveis, ao rolar a tela para baixo uma nova linha aparece) ou por páginas (o conteúdo é separado por páginas, todos os caracteres são visíveis e ao rolar a tela para baixo uma nova página aparece). Para avaliar a usabilidade de cada um desses métodos foram selecionados 15 usuários, todos com experiência no uso de telefone celular, mas nenhum com experiência no uso de aplicações WAP. Utilizando o emulador foi solicitado a cada um dos participantes que executasse três tarefas diferentes, uma para cada método, após a realização das quais eles deveriam responder a seis questões sobre o texto que haviam lido na tela. Foram então medidos o tempo que o participante levou para responder às seis questões e o número de respostas erradas. Após experimentar os três métodos de apresentação de texto foi solicitado ao participante que avaliasse o método atribuindo um valor em uma escala de usabilidade. As sessões foram registradas em vídeo e os resultados apontaram o método de rolagem vertical como o melhor entre as três opções.

Embora o emulador seja uma forma prática, rápida e praticamente sem custos para testar o código da aplicação, ele apresenta algumas limitações quanto à sua utilidade para avaliar a usabilidade (Pearrow, 2002). Uma das principais considerações diz respeito às limitações ergonômicas que devem ser consideradas. Ao usar um emulador, o usuário interage com um equipamento virtual vendo a tela do computador e fazendo a entrada de dados através do mouse e do teclado ou diretamente na tela, caso ela seja sensível ao toque (figura 28). Ao utilizar o emulador é possível que se percam informações importantes sobre a usabilidade da interface que só apareceriam quando o usuário estivesse interagindo fisicamente com o equipamento real (Weiss, 2002).



Figura 28: Usuário utilizando um emulador no computador

Fonte: (Kaasinen, 2000)

Segundo Chittaro (2001) ao realizar o teste utilizando um emulador no computador:

A experiência sensorial do usuário é completamente diferente do ponto de vista visual e auditivo, os periféricos de entrada/saída e os estilos de interação são completamente diferentes (interação com o ponteiro do mouse em controles virtuais X interação direta com controles físicos).

O estudo de Vyas (2002) teve como objetivo gerar um conjunto de *guidelines* para o desenvolvimento de aplicações WAP em telefones celulares. O autor utilizou um emulador para avaliar a usabilidade de dois sites WAP e verificar a aplicabilidade das *guidelines*. Foram selecionados quatro participantes, todos com experiência no uso de telefone celular mas nenhum havia usado uma aplicação WAP anteriormente. Os participantes deveriam navegar nos dois sites escolhidos e executar seis tarefas como ler *e-mail*, encontrar a programação de cinema ou as principais manchetes com as últimas notícias. Foi medido o tempo que o participante levou para executar cada tarefa e a maneira como ele a executou. Posteriormente, as ações do usuário foram interpretadas para inferir como essas ações estavam relacionadas aos problemas de usabilidade no *design* da interface. O participante era encorajado a comentar todos os problemas de interação durante a realização da tarefa e a sua

opinião sobre o site, o que ele gostou e o que não gostou. Após o teste o participante foi entrevistado para poder criticar e comentar sobre o *layout* e a apresentação dos sites. Justificando que seu objetivo era somente avaliar a aplicabilidade das *guidelines*, o autor considerou apropriado utilizar um emulador no computador para a avaliação de usabilidade. No entanto ele ressalta em seu estudo (Vyas, 2002, p.50) que:

... obviamente a avaliação feita com um emulador WAP não dará os mesmos resultados, em termos de usabilidade, que a avaliação realizada com um telefone WAP. Na avaliação com emuladores aspectos como ergonomia física, estilos de interação, mecanismos de entrada de dados, etc. não serão aplicáveis uma vez que no telefone real eles se dão de forma completamente diferente.

Se o produto for novo ou se não existir um emulador adequado ao equipamento, Weiss (2002) sugere a construção de protótipos usando programas de autoria multimídia como Toolbook, Hypercard ou Flash em páginas Web. Com estas ferramentas é possível criar várias imagens que simulam a aparência final do produto e definir algumas regiões da tela que podem ser clicadas simulando o seu funcionamento. Entretanto, como esta abordagem é mais limitada, uma vez que ela não proporciona todas as funcionalidades de um emulador, ela está mais indicada para demonstrações, simples testes de *layout* da aplicação ou como uma alternativa ao protótipo em papel na técnica de Prototipação em Papel (Nielsen, 2003a).

Bocker (1999) utilizou esta abordagem para avaliar a interface com o usuário durante o projeto de um novo modelo de telefone celular para o qual o equipamento real ainda não existia. Foram construídos dois protótipos multimídia no computador simulando dois modelos de interfaces, uma com ícones e outra somente com texto, e estas foram avaliadas para verificar qual a preferência dos usuários e em qual delas eles obtinham melhores índices de desempenho. Ao todo 80 pessoas participaram das avaliações. Foram escolhidos participantes com e sem experiência

tanto no uso de telefones celulares quanto no uso de Windows. Os 80 participantes foram divididos em 2 grupos, 40 testaram a interface com ícones e 40 testaram a interface com texto. Metade dos participantes leu o manual do usuário antes de executar as tarefas e a outra metade não leu, mas poderia consultá-lo durante a realização das tarefas. Com base nestes critérios foram estabelecidos 4 grupos de 20 participantes: ícones X texto e leitura do manual obrigatória X leitura do manual opcional. Foram criados cenários com tarefas simples como fazer uma ligação, fazer uma ligação para um número de telefone gravado na agenda, procurar por um número de telefone na agenda durante uma ligação, gravar um número de telefone para acesso rápido. No final das tarefas os participantes responderam a questionários e entrevistas. As sessões duraram uma hora e meia e foram realizadas por dois avaliadores que estavam na mesma sala que o participante, um avaliador auxiliava o participante e coletava dados a partir da observação do usuário e o outro gravava as perguntas e comentários do participante e cronometrava o tempo. As sessões foram gravadas em vídeo com imagens da tela do computador e da face do usuário. Na análise das sessões o avaliador verificava se as tarefas foram cumpridas, quanto tempo o participante levou para executá-las, quantos erros foram cometidos e quantas vezes o participante consultou o manual do usuário. Bocker relata que "... a avaliação revelou os problemas mais cruciais de usabilidade identificando também quais os pontos fortes do design da interface". Para o autor "os resultados da avaliação foram extremamente satisfatórios e serviram de base para importantes decisões sobre as direções do projeto".

Apesar de todas as limitações, utilizar um emulador no computador pode ser a única opção quando não é possível ter acesso ao equipamento real que deve ser testado. Além de ser possível utilizar um emulador para cada uma das plataformas

que será avaliada, cada integrante da equipe de desenvolvimento pode ter uma cópia do emulador em seu próprio computador, diferentemente do equipamento real que precisaria ser compartilhado entre todos. Outro ponto a ser considerado é em relação à rápida evolução dos equipamentos, pois qualquer atualização que seja feita no equipamento real poderá ser reproduzida no emulador rapidamente e praticamente sem custos o que nem sempre é possível de realizar nos equipamentos de teste (Pearrow, 2002).

3.2 Avaliação de usabilidade com o equipamento real

Chittaro (2001) optou por utilizar um telefone WAP real em seus testes para avaliar diferentes estilos de navegação em sites WAP por “não acreditar que estudos sobre sites WAP utilizando emuladores no computador pudessem fornecer indicações confiáveis da performance do usuário”.

Quando a avaliação de usabilidade é realizada com o próprio equipamento e se deseja gravar em vídeo a interação, um recurso que pode ser empregado é a câmera de documentos (figura 29). A câmera é colocada sobre uma mesa com o equipamento montado sobre um tripé, o usuário deve estar sentado e o aparelho não pode ser movimentado para que seja possível filmar com boa qualidade tanto a tela quanto o teclado do aparelho (Pearrow, 2002; Weiss, 2002). É possível utilizar também microfones e outras câmeras de vídeo para registrar as reações do usuário. Estes sinais podem ser combinados gerando uma fita de vídeo onde a tela está dividida em duas metades: na primeira metade aparecem as imagens da câmera de documentos, que mostram a interação com o equipamento, e na outra metade estão as imagens da câmera de vídeo, que mostram a face do usuário.



(a) câmera ELMO

(a) Fonte: (ELMO, 2002)

© Copyright 2000-2002 ELMO.

Todos os direitos reservados



(b) câmera PCDC-3

(b) Fonte(USI, 2002)

© Copyright 2002. Usability

Systems, Inc.

Todos os direitos reservados



(c)usuário realizando a avaliação

(c) Fonte: (Davies, 2002)

© Copyright 2002. 3G Lab.

Todos os direitos reservados

Weiss (2001) utilizou uma câmera de documentos em seu estudo para avaliar a usabilidade de um telefone celular e dos serviços disponíveis pela operadora de telefonia móvel. Foram escolhidos 6 participantes, com idades entre 21 e 44 anos, variando de inexperientes até usuários freqüentes da Web. Todos eram usuários de telefone celular e nenhum tinha profissões relacionadas às áreas de Internet, informática, telefonia ou marketing. Os testes foram realizados em uma sala normal que não possuía espelho falso com dois avaliadores presentes. Um avaliador sentava-se ao lado do participante e pedia que este realizasse oito tarefas simples como verificar a previsão do tempo, checar horários de cinemas, ler e enviar *e-mails*. O outro avaliador observava as ações do participante, verificava se ele tinha sucesso ou não na realização da tarefa, cronometrava o tempo que ele levava para executá-

la e media a sua taxa de erros. Os testes duraram uma hora e foram registrados em vídeo e em áudio. O telefone foi fixado com velcro sobre um tripé colocado em cima da mesa para permitir que o participante segurasse o telefone com as mãos. Foi usada uma câmera de documentos para gravar as imagens da tela e do teclado do telefone e uma câmera de vídeo normal para filmar a face do participante. Ambas as imagens foram colocadas na tela lado a lado para análise simultânea da imagem do telefone e da face do participante.

Em seu estudo para comparar a usabilidade de dois modelos de telefones celulares com câmera fotográfica Davies (2002) também utilizou uma câmera de documentos. Os objetivos eram avaliar a função de máquina fotográfica dos equipamentos e realizar uma entrevista com os participantes para verificar qual dos dois telefones o usuário escolheria para comprar e qual motivo o levaria a escolher um ou outro. Seis usuários, entre 22 e 34 anos, entre os quais três eram especialistas e três eram novatos no uso do telefone celular, participaram das avaliações. O participante escolhia um dos dois modelos para usar primeiro e deveria então tirar uma foto, armazenar, renomear, visualizar e enviar essa foto por *e-mail*. Depois ele deveria realizar as mesmas tarefas com o segundo modelo de telefone celular. No final o moderador pedia que o participante comparasse os dois modelos e escolhesse qual deles ele compraria. A avaliação durou uma hora e foi realizada no laboratório com o participante e o facilitador em uma sala e os observadores em outra, separados pelo espelho falso. O telefone foi montado em um tripé sobre a mesa com uma câmera de documentos filmando o teclado e a tela do aparelho. Uma outra câmera de vídeo filmava a face do usuário. Os participantes eram encorajados a falar o que estavam fazendo e todo o áudio da sessão também foi registrado.

Alguns autores preferem não utilizar o equipamento montado sobre o tripé mas deixar que o usuário o segure nas mãos. Segundo Mizobuchi (2002) a posição do usuário e a maneira como ele segura o equipamento são fatores que afetam o uso da interface e também devem ser considerados. Para o autor:

A performance do usuário vai ser afetada dependendo da mão que ele utiliza para acionar o equipamento e também de sua postura. É diferente o usuário operar o equipamento quando está sentado em uma cadeira ou quando está parado em pé no meio da rua.

Em sua pesquisa sobre a usabilidade do protótipo de uma aplicação WAP para telefones celulares Chincholle (2002b) realizou os testes dentro do laboratório com o usuário sentado em uma poltrona, segurando o telefone nas mãos enquanto uma câmera colocada no teto filmava a tela e o teclado do telefone. A aplicação avaliada era um serviço que oferecia mapas, escolha de caminhos e determinados conteúdos que dependiam da localização do usuário (onde ele se encontrava e onde ele desejava ir), como por exemplo: encontrar o melhor caminho para ir de carro até uma determinada rua ou encontrar um restaurante nas proximidades e descobrir qual o melhor caminho para chegar até lá a pé. Participaram da avaliação sete usuários, entre 26 e 59 anos, todos com experiência no uso do modelo de telefone que estava sendo usado no experimento, mas sem conhecimento da aplicação avaliada, quatro já haviam usado PDA e cinco já haviam usado outros modelos de telefone celular. O usuário ficava sozinho na sala, separado do observador pelo espelho falso e era filmado por várias câmeras de vídeo. Ele deveria executar cinco tarefas e enquanto interagia com a aplicação deveria comentar sobre tudo o que estava fazendo e sobre os problemas que estava enfrentando. O avaliador verificava somente se a tarefa havia sido concluída sem cronometrar os tempos de execução das tarefas. O critério adotado pelo autor para determinar o que ele denominou como “acurácia” foi baseado em como o participante realizou a tarefa em relação à

necessidade de ajuda: se ele completou a tarefa sozinho a acurácia foi considerada 1, se recebeu alguma ajuda a acurácia foi 0.5 e se precisou ser guiado pelo avaliador a acurácia foi zero. Ao final das tarefas os participantes responderam a um questionário de satisfação para avaliar a sua opinião sobre a aplicação.

McIntyre (2002) relata sua experiência na realização de um teste remoto para avaliação de usabilidade de um aplicativo i-mode, acessado por PDA's ou telefones celulares, que precisava ser realizada no Japão mas com o avaliador nos EUA. A solução encontrada foi utilizar uma pessoa que não tinha experiência na condução de testes de usabilidade para realizar as avaliações no Japão e gravar toda a interação em fitas de vídeo que foram enviadas para o avaliador para análise. O avaliador desenvolveu um conjunto de *guidelines* a serem seguidas pelo facilitador do teste para que, mesmo sem experiência, ele pudesse conduzir as avaliações sem comprometer os resultados. Seis participantes utilizaram seus próprios equipamentos para executar quatro tarefas e no final responderam a um questionário sobre a sua opinião em relação ao aplicativo. Foi utilizada uma filmadora comum para gravar as imagens da tela e do teclado do equipamento enquanto o participante interagia normalmente com a aplicação. As fitas de vídeo foram enviadas ao avaliador que pôde analisá-las para identificar: se o participante havia completado ou não a tarefa, qual havia sido a duração e a frequência das pausas dos usuários, quais as dúvidas dos participantes em relação aos próximos passos durante a execução das tarefas, e quais os comentários positivos e negativos. McIntyre (2002) considerou o resultado da avaliação como positivo em função da grande quantidade de problemas de usabilidade que esta foi capaz de identificar.

O estudo de Chittaro (2001) avaliou diferentes formas de implementação de estilos de navegação em uma aplicação WAP para telefone celular. O autor selecionou quarenta pessoas que segundo ele deveriam ser representativas do público em geral e, portanto, variavam quanto a idade, ocupação e habilidade no uso do computador, mas nenhuma havia usado um telefone WAP anteriormente. Os testes foram realizados na casa do participante para que ele se sentisse mais à vontade. Uma câmera de vídeo foi utilizada para registrar a interação, mas o participante não foi filmado. O participante deveria executar duas tarefas: encontrar alguns filmes que estivessem passando no cinema e reservar lugares para um horário escolhido. As medidas de performance foram definidas como: tempo total para executar a tarefa, número de passos para completar a tarefa, percentual da tarefa que o participante conseguiu completar em 9 minutos (um teste piloto feito previamente determinou que 9 minutos seria o tempo médio para a execução da tarefa) o número de desvios em relação ao caminho ótimo para realizar a tarefa e o número total de *cards* que foram visitados (*cards* equivalem a páginas Web). Ao final do teste, uma avaliação subjetiva das interfaces foi realizada onde cada participante respondeu a um questionário sobre a sua opinião a respeito da interface.

O objetivo de Rao(2000) foi comparar três métodos distintos de entrada de dados em PDAs (a caneta, o teclado e ativação por voz) para: determinar qual é o método de entrada de dados mais eficiente em termos de velocidade e precisão, determinar qual é o método, ou a combinação de métodos, preferido pela maioria dos usuários e elaborar um conjunto de guias de estilo para auxiliar no projeto de aplicações para PDAs. Foram recrutadas treze pessoas, entre 22 e 33 anos, todos pertencentes aos departamentos de Ergonomia e Psicologia da universidade, que utilizaram três modelos diferentes de PDA's. O participante deveria executar tarefas comuns como

salvar ou editar um endereço, utilizando as diferentes formas de entrada de dados disponíveis nos PDAs. O teste foi realizado dentro do laboratório com o usuário sendo filmado para verificar em quanto tempo ele executava a tarefa e com que precisão. Na seqüência cada participante era entrevistado para falar sobre as suas atitudes e opiniões em relação a cada método de entrada de dados que havia utilizado.

O trabalho de (Öquist,2002) tinha como objetivo comparar diferentes formas de apresentação de texto em um PDA. Participaram deste estudo dezesseis pessoas, com uma média de idade de 25 anos, todos com experiência prévia no uso de computador e sete com experiência prévia no uso de PDA. Os usuários fizeram os testes em um laboratório com recursos de áudio e vídeo. O usuário ficava sozinho na sala de testes e o observador o acompanhava em uma outra sala atrás do espelho falso. Cada participante deveria ler dois textos, um curto e um longo. Ao final o participante deveria preencher dois questionários, o primeiro sobre o conteúdo do texto lido para avaliar qual foi a sua compreensão e o segundo para avaliar o esforço cognitivo empreendido na leitura. Foi também medido o tempo que o usuário levava para ler cada um dos textos.

Além das câmeras de vídeo alguns autores utilizam ferramentas de software que registram as interações do usuário no próprio equipamento através de arquivos gerados durante a utilização do sistema. Alguns desses estudos estão descritos a seguir:

O trabalho de Ericsson (2001) teve como objetivo verificar se o aparelho de telefone celular e se o *browser* têm influência na usabilidade do serviço WAP. Dez participantes, com idades entre 22 e 61 anos participaram das avaliações com duração de uma hora e meia. Todos os participantes tinham experiência no uso da

Internet e de vários modelos de telefones celulares, mas nenhum tinha experiência no uso de WAP. Foram utilizados na avaliação dois modelos de telefone celular com dois *browsers* completamente diferentes. Três diferentes serviços WAP foram usados em três tarefas, sendo que todos os participantes executaram as três tarefas com os dois modelos de telefones. O teste foi feito no laboratório e gravado em áudio e vídeo. A acurácia da tarefa foi medida utilizando uma ferramenta para análise dos dados, chamada KeystrokeMapper (Nyberg, 2001), que comparava o caminho feito pelo usuário com o caminho ótimo para a realização da tarefa. Também foi medido o tempo total de execução da tarefa.

Uma ferramenta que registrava toda a interação no PDA foi utilizada por (Mizobuchi,2002) com o objetivo de comparar a performance do usuário na utilização de diferentes métodos de entrada de dados como o uso da caneta (*stylus*) ou das teclas para controle do cursor. Dez pessoas, com idades entre 19 e 60 anos, e que nunca haviam utilizado um PDA anteriormente participaram dos testes. As avaliações foram realizadas dentro do laboratório, e o participante segurava o PDA em suas mãos. O registro da interação no próprio equipamento permitiu medir o tempo que o usuário levava para clicar em diversos alvos apresentados na tela. Puderam então ser comparados os tempos para clicar no alvo com a caneta e com as teclas para controle do cursor. Também foi possível verificar qual o efeito do tamanho do alvo e do número de alvos presentes na tela na performance do usuário. Os resultados mostraram que o uso da caneta provocou mais erros para alvos menores. Ao aumentar o tamanho dos alvos os dois métodos apresentaram a mesma taxa de erros.

Sears (2003) utilizou ferramentas de análise em seu estudo comparativo para avaliar qual o impacto do tamanho dos teclados dos PDA's na performance e na

preferência dos usuários. Trinta participantes foram divididos em três grupos de dez pessoas e cada um dos grupos utilizou um tamanho diferente de teclado (pequeno, médio e grande). Os participantes tinham idade média de 26 anos, todos estudantes universitários, mas não de cursos relacionados à Tecnologia da Informação ou a Engenharias. Todos tinham experiência no uso do computador, mais da metade não era usuário de telefone celular e nenhum deles havia usado um PDA anteriormente. Os participantes podiam optar entre apoiar o PDA sobre a mesa ou segurá-lo nas mãos. Eles deveriam executar seis tarefas, todas relacionadas somente à digitação de texto, e toda a interação foi registrada no próprio equipamento. A performance do usuário foi medida em número de palavras por minuto e em número de erros. No final os participantes responderam a um questionário de satisfação em relação ao sistema que avaliava itens como facilidade, rapidez e precisão na digitação de texto. Curiosamente, neste estudo o tamanho dos teclados não teve influência nas medidas de performance avaliadas, revelando também que a diminuição do tamanho do teclado não aumentou a taxa de erros ou impactou negativamente na preferência dos participantes.

Algumas outras avaliações não utilizam câmeras para filmar a interação e são baseadas na observação direta do usuário e em questionários e entrevistas, como as descritas a seguir:

O estudo de Helyar (2000) teve como objetivo avaliar a reação e a opinião das pessoas em relação aos serviços WAP. Durante uma hora e meia doze participantes utilizaram vários serviços WAP executando tarefas comuns como ler as principais notícias e verificar a programação de cinema. As avaliações foram conduzidas dentro do laboratório com o avaliador e o participante em uma sala e os observadores em outra sala, separadas por um espelho falso. Enquanto navegavam

por vários serviços WAP os participantes explicavam ao avaliador o que estavam fazendo e quais eram as suas dúvidas e inquietações. O autor denominou esta avaliação de colaborativa uma vez que todos os resultados foram baseados na conversa entre o avaliador e o participante enquanto este interagia com o sistema.

O estudo de Duda(2001) teve como objetivo investigar o porquê da baixa aceitação dos serviços WAP e para identificar quais os principais problemas enfrentados pelo usuário no uso desses sistemas. Trinta e seis participantes foram selecionados, com idade média de 29 anos, todos com experiência no uso de telefones celulares. Foram escolhidos três modelos diferentes de telefones celulares, quatro portais WAP de diferentes operadoras e um total de 23 serviços WAP variados como cinema, ações, livrarias, guias de lazer, serviços sobre trânsito, ferramentas de busca e serviços sobre saúde. Os participantes recebiam os telefones, os portais e os provedores de forma randômica. O participante deveria realizar doze tarefas ao final das quais ele respondia a um questionário e a uma entrevista feita pelo avaliador sobre o serviço que ele acabara de utilizar. No final do teste era feita uma última entrevista geral. Ao observar o comportamento do usuário, os seus comentários e as suas respostas aos questionários e às entrevistas o avaliador pôde quantificar a aceitabilidade, a utilidade e a usabilidade dos serviços testados.

3.3 Avaliação de usabilidade fora do laboratório

Segundo Brewster(2002, p.203) “é de vital importância que os computadores de mão sejam testados em situações reais para verificar a usabilidade de suas

interfaces". Os estudos a seguir foram realizados fora do laboratório procurando colocar o usuário em uma situação mais próxima ao contexto real de uso.

Waterson (2002) propõe a utilização de ferramentas de software específicas para a realização de testes remotos com usuários móveis. A autora analisa dois tipos de avaliação de usabilidade para acesso à Internet com PDAs através da comparação feita entre os dados coletados no teste tradicional no laboratório e no teste remoto. Dez participantes utilizaram um PDA para navegar em um site WAP com o objetivo de encontrar informações sobre automóveis. Metade dos participantes fez o teste no laboratório da forma tradicional com o PDA sobre a mesa e com câmeras gravando, em áudio e vídeo, a interação do usuário com o equipamento. Um avaliador estava presente para registrar comentários e para responder a qualquer questão do participante. Os outros cinco participantes receberam o PDA e foram liberados para fazer as tarefas solicitadas onde e quando eles desejassem sem a presença do avaliador e sem câmeras. Para estes foi usado o teste remoto com a ferramenta WebQuilt para análise de *logs*. Através de uma interface para visualização dos dados coletados era possível obter informações sobre: quais páginas Web foram visitadas, onde cada participante clicou em cada página e quais os caminhos mais utilizados pelos participantes para realizar a tarefa. As instruções sobre as tarefas, os dados sobre os participantes e o questionário respondido ao final do teste foram apresentados ao usuário no próprio PDA. A análise de *logs* foi feita para os dez participantes do teste.

Em seu estudo sobre a usabilidade do WAP, Nielsen (2000) entregou um telefone para cada usuário utilizar durante uma semana e pediu que eles anotassem suas experiências em um diário. O autor classificou essa metodologia de mista e justificou sua escolha porque (Nielsen, 2000, p.12):

...tanto o WAP quanto a Internet móvel em geral são inerentemente pessoais, portanto uma grande parte do estudo dependeu de como o usuário utilizou o equipamento enquanto estava sozinho...Entretanto o valor de observar o comportamento do usuário em condições controladas era inestimável, pois permitia ter uma visão mais aprofundada sobre o impacto da interface com o usuário no uso do WAP.

Foram ao todo dez participantes, com idades entre 19 e 50 anos, usuários de Internet variando de inexperientes até usuários freqüentes, todos usavam telefones celulares, mas nunca haviam utilizado WAP, apesar de ter interesse nesse serviço. O participante deveria executar tarefas comuns como verificar a previsão do tempo, encontrar o telefone de um determinado restaurante, verificar a programação da TV ou procurar pelo resultado de algum jogo do dia anterior. Os participantes receberam um telefone para usar durante uma semana e deveriam anotar em um diário todas as suas impressões durante o uso do WAP. Foram orientados a usar o telefone todos os dias da semana de 30 min a 1 hora em pequenos períodos espalhados ao longo do dia. Foram preenchidos dois questionários, um antes de começar os testes e outro no final da semana que eles passaram com os telefones. Na observação direta, feita no início e no final da semana, o avaliador cronometrou o tempo que o participante levava para executar um conjunto de tarefas para poder avaliar se a eficiência do participante aumentou após a semana de uso e também para verificar se ele estava mais familiarizado com o uso do WAP. Ao final da semana foi feita uma entrevista, gravada em áudio, com os participantes para que eles pudessem relatar suas experiências. Nesta entrevista o participante era encorajado a recorrer ao diário para verificar suas anotações e lembrar o que aconteceu.

Deixar o participante explorar o equipamento e suas aplicações por conta própria durante uma semana parece uma abordagem interessante quando se trata de um equipamento de uso pessoal como o telefone celular. Entretanto pode ser custoso

em termos de tempo e recursos financeiros, pois para testar com vinte usuários Nielsen (2000) precisou utilizar quatro telefones para que o teste pudesse ser realizado em cinco semanas. Na maioria das vezes não se tem tanto tempo disponível e tantos recursos para realizar uma única avaliação de usabilidade. Outro ponto importante se refere, principalmente, à boa vontade e à disciplina exigidas do participante para utilizar o equipamento e relatar com precisão todos os problemas e dificuldades encontrados. Nielsen reconhece que a aplicação desta técnica foi possível porque um dos avaliadores possuía vasta experiência na realização de documentários para a televisão e portanto tinha amplo conhecimento sobre a utilização de técnicas para coleta de testemunhos da forma mais natural, imparcial e precisa possível. Esta é uma consideração importante uma vez que grande parte da informação sobre os problemas de usabilidade encontrados foram colhidos a partir do relato dos participantes.

Os estudos de Petrie (1998) sobre interfaces para computadores de mão para auxiliar a navegação de pessoas cegas foram realizados fora do laboratório pois seu objetivo era obter medidas de usabilidade em situações mais reais de uso. Para o autor "...as metodologias para avaliação de interfaces dos computadores de mesa devem ser adaptadas e complementadas para que possam detectar as peculiaridades dos computadores de mão". O objetivo deste estudo foi avaliar a usabilidade e a aceitação do sistema MoODS (MoBIC Outdoor System), um "*wearable system*" projetado para auxiliar a movimentação de viajantes cegos. As pessoas cegas têm dois tipos principais de obstáculos a enfrentar, os da chamada micro-navegação, que são os obstáculos que estão mais próximos e os da macro-navegação que são as orientações referentes ao ambiente em que ela está, como a rua onde se encontra, em que direção ela está se deslocando, quando cruzar o sinal

com segurança, etc. A micro-navegação pode ser auxiliada por uma bengala ou por um cão-guia. A macro-navegação, difícil em ambientes conhecidos, se torna praticamente impossível em ambientes desconhecidos. Para auxiliar este deslocamento na macro-navegação foi desenvolvido um protótipo, que utiliza GPS, para localizar o usuário em relação à rua, aos números das casas, aos pontos turísticos próximos e também às informações de interesse específico para os viajantes cegos. A interação com o sistema pode ser ativada pelo usuário, quando ele não tem certeza de sua localização, ou pelo sistema, quando o próprio sistema manda algum sinal de aviso ao usuário. O sistema é composto por um fone de ouvido para recebimento das mensagens pré-gravadas (colocado na frente das orelhas e não sobre elas), uma pulseira com oito teclas para a entrada de dados e a caixa com as baterias que pode ser levada nos bolsos ou em uma sacola. Participaram do experimento sete pessoas, com idades entre 18 e 65 anos, todas cegas. Os participantes estudavam dois trajetos desconhecidos utilizando o MoBIC Indoor System (sistema de voz para estudo de mapas e planejamento de rotas). Quando estivessem seguros, eles deveriam sair e percorrer esses trajetos. Na primeira vez eles faziam uso do MoODS e da sua bengala ou cão-guia. Na segunda vez o trajeto era percorrido sem o MoODS, somente com a bengala ou cão-guia. Cada trajeto tinha aproximadamente 1 km e os dois trajetos apresentavam o mesmo nível de dificuldade. A performance do usuário foi medida objetivamente através do tempo necessário para percorrer o trajeto, número de erros cometidos, número de vezes que o MoODS foi acionado. Foi também medido o índice PPWS-Percentage Preferred Walking Speed para avaliar a velocidade com que o participante caminhava. Toda pessoa tem uma velocidade padrão para andar que parece ser a mais eficiente fisicamente para cada um, qualquer objeto que a pessoa vá utilizar

estando em movimento, não deve afetar esta velocidade. Avaliar a velocidade com que a pessoa anda através deste índice pode indicar o grau de stress a que ela está submetida. Foram medidos os batimentos cardíacos e todos os participantes responderam a questionários para avaliar seu nível de stress e ansiedade, esforço cognitivo e a sua satisfação em relação ao sistema.

Para Brewster (2002) é fundamental fazer testes de usabilidade das interfaces para equipamentos móveis fora do laboratório, com o usuário realizando outras atividades em paralelo. Seu estudo tinha como objetivo avaliar se a adição de som ao clicar um botão na tela poderia auxiliar a entrada de dados em telas de toque pequenas aumentando a usabilidade e permitindo que os tamanhos dos botões pudessem ser reduzidos. O autor utilizou um PDA para testar várias versões de interface com diferentes tamanhos de botões e com o uso ou não de som associado a cada botão. Os testes foram realizados dentro do laboratório e também ao ar livre. Em cada uma das situações de teste participaram dezesseis pessoas, com idades entre 21 e 45 anos, estudantes e funcionários do departamento de Ciência de Computação da universidade. Todos conheciam interfaces gráficas, alguns já haviam usado algum modelo de PDA ou telefone celular mas nenhum conhecia a interface do Palm. O participante deveria digitar valores apontando a caneta no teclado numérico que aparecia na tela. O número de 5 dígitos que deveria ser digitado aparecia no canto direito da tela, o participante deveria digitar os 5 dígitos, pressionar OK e continuar com o próximo número. Os números eram gerados aleatoriamente pela própria aplicação que também armazenava o número total de cliques na tela. Foram medidos o número de dígitos que o participante conseguiu clicar e o esforço cognitivo através de um questionário. Também foi perguntado ao participante com qual das duas interfaces ele achou mais fácil realizar a tarefa. O

primeiro experimento foi realizado com os participantes sentados, testando as interfaces com e sem som dentro do laboratório. No segundo experimento eles realizaram as mesmas tarefas enquanto andavam por uma determinada rua dentro do campus da universidade. O percurso era uma rua reta com aproximadamente 10 metros de comprimento na qual o participante deveria ir e voltar várias vezes até que o seu tempo acabasse. O primeiro experimento foi em um ambiente controlado, silencioso com o participante parado podendo se concentrar na tarefa. O segundo experimento foi feito ao ar livre com o usuário andando enquanto entrava com os dados. Neste segundo caso foi contado também o número de voltas que o participante completava. De acordo com os resultados deste estudo a usabilidade foi afetada (maior carga cognitiva e uma redução de 32% na entrada de dados) quando o equipamento foi usado fora do laboratório. Para o autor este resultado sugere que utilizar uma aplicação de entrada de dados em um computador de mão em um ambiente mais próximo ao real pode afetar significativamente a interação. Brewster (2002, p.191) afirma que:

É importante testar ao ar livre, em situações mais reais, com o usuário tentando utilizar a interface enquanto executa outras tarefas, como andar na rua. Os equipamentos móveis não devem ser testados somente dentro do laboratório onde muitos problemas importantes com a interface podem passar despercebidos.

3.4 Minicâmeras para avaliação de usabilidade de computadores de mão

Para Nyssonnen (2002) "ao executar testes de usabilidade com aplicações móveis em equipamentos portáteis as pessoas gostariam de poder segurá-los nas mãos e mover-se livremente como elas fariam normalmente". Entretanto, quando é necessário filmar a interação, a utilização de câmeras comuns para filmar telas pequenas em movimento pode não trazer resultados muito precisos. O autor

apresenta uma solução composta de uma minicâmera acoplada diretamente ao telefone celular e ligada por um fio até uma filmadora que grava a imagem que está sendo transmitida. Um pequeno microfone também conectado por fio a um amplificador e à filmadora é utilizado para gravar o áudio. O usuário pode mover as mãos livremente dentro dos limites dos fios que ligam a câmera e o microfone à filmadora.

Para que o participante possa ter mais liberdade de movimentos pode ser utilizada uma minicâmera sem fio como a apresentada por (Pers, 2002). Esta solução permite filmar a interação enquanto o usuário se movimenta à vontade, e o avaliador o observa a distância. A câmera é acoplada ao computador de mão, possibilitando desta forma registrar as imagens da tela e do teclado. Os sinais de vídeo da câmera são enviados a um receptor que pode ser conectado a uma filmadora, a um monitor de vídeo ou a um computador. A distância entre o usuário e o avaliador depende do modelo da câmera que está sendo utilizada, e para esta solução pode chegar a 25 metros. A figura 30 ilustra um possível teste realizado fora do laboratório utilizando uma minicâmera sem fio.

Estas duas minicâmeras ilustram o desenvolvimento de novas tecnologias adaptadas para a avaliação de usabilidade dos computadores de mão, entretanto não há até o momento nenhum estudo publicado sobre a realização de avaliações de usabilidade que tenham feito uso destes equipamentos. Nyysönen (2002) e Pers (2002) somente descrevem estas soluções tecnicamente.



Figura 30: Ensaio de interação utilizando uma minicâmera sem fio

3.5 Prática nas empresas

Para avaliar como as avaliações de usabilidade para computadores de mão estão sendo conduzidas na prática pelas empresas, foi realizada no mês de novembro de 2002 uma pesquisa com os participantes de um grupo de discussão da Internet (Usablemobile, 2002), um dos principais fóruns para troca de informações sobre usabilidade de interfaces para equipamentos móveis. Esta pesquisa teve por objetivo verificar quais ferramentas estão sendo utilizadas nos testes de usabilidade e com que frequência as empresas realizam os testes fora do laboratório.

A pesquisa foi enviada por *e-mail* a todos os participantes do grupo Usablemobile. Doze pessoas, representando sete países (figura 31), responderam a esta pesquisa. Todas as respostas vieram de profissionais envolvidos com a avaliação de usabilidade para computadores de mão, pertencentes a grandes empresas e centros de pesquisa em computação móvel no mundo.

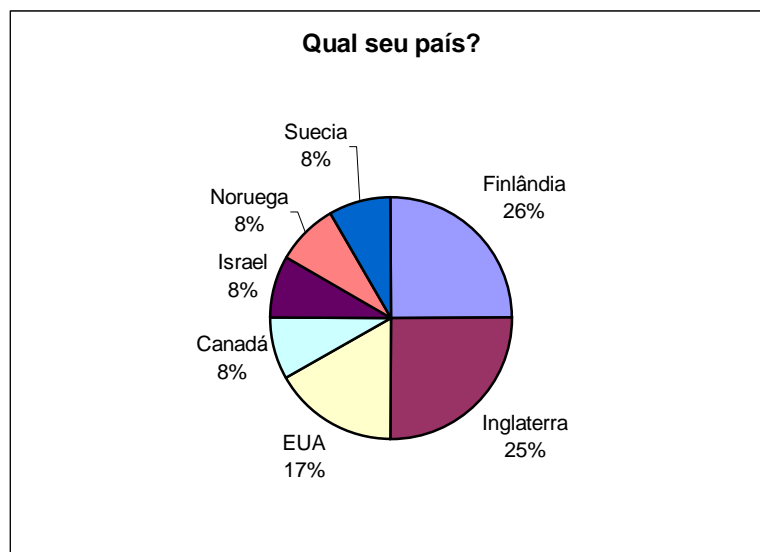


Figura 31: Países que responderam à pesquisa

As respostas sobre os equipamentos utilizados indicam que tanto os emuladores quanto as câmeras de documentos representam ferramentas importantes nas avaliações de usabilidade, sendo usados respectivamente, por 80% e 70% dos profissionais (figura 32).

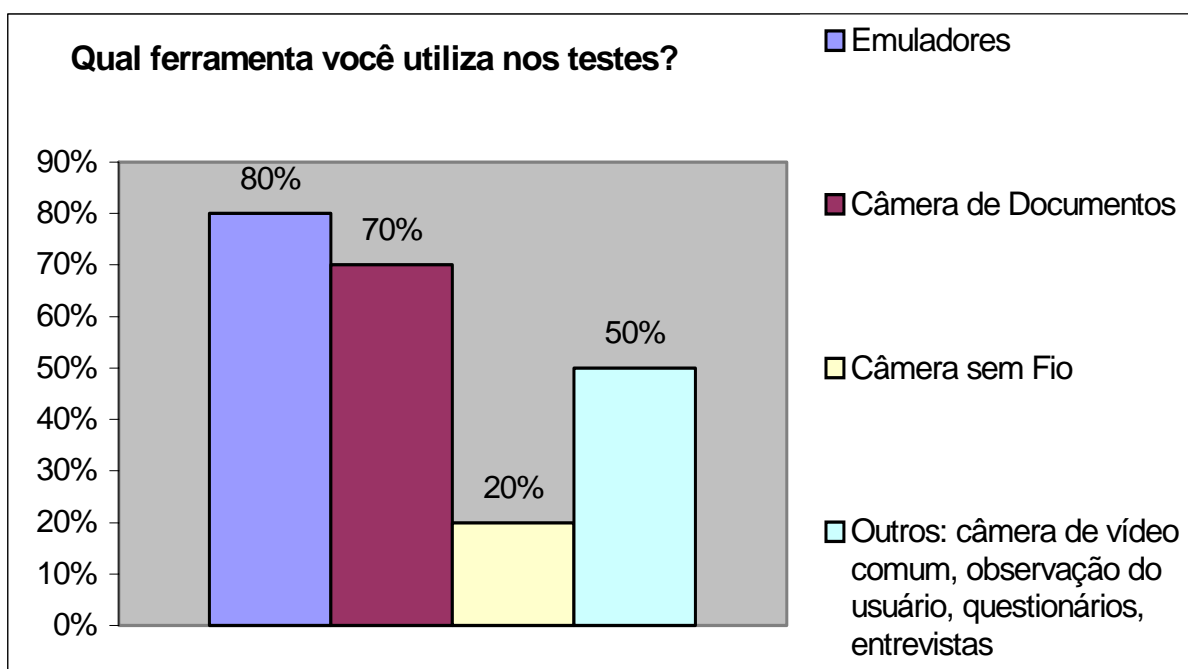


Figura 32: Ferramentas utilizadas nos testes de usabilidade

Quanto ao local dos testes, 50 % dos profissionais dizem realizar testes fora do laboratório (figura 33). Nestes casos a técnica mais freqüente é a observação do usuário, muitas vezes utilizando uma câmera de vídeo para capturar imagens de situações reais de uso.



Figura 33: Local dos testes de usabilidade

Apesar da pequena amostra esta pesquisa pode ser considerada representativa, uma vez que os respondentes são profissionais de grandes empresas da área de computação móvel, todos eles responsáveis pela avaliação de usabilidade de equipamentos e aplicações.

3.6 Escolha das três abordagens para os ensaios de interação

Este capítulo ilustrou como vêm sendo conduzidas as avaliações de usabilidade para os computadores de mão, seja pelos pesquisadores ou por algumas das principais empresas da área.

A partir dessa revisão bibliográfica e dos resultados da pesquisa realizada com os profissionais da área, foi possível identificar três principais abordagens para a

realização de ensaios de interação para avaliar a usabilidade dos computadores de mão baseadas no uso do emulador, da câmera de documentos e da minicâmera sem fio.

Esta escolha se deu a partir da constatação do expressivo uso do emulador e da câmera de documentos como métodos para registrar a interação com os computadores de mão dentro do laboratório enquanto que a minicâmera sem fio foi a solução encontrada para registrar a interação na realização dos testes fora do laboratório.

O capítulo a seguir relata o experimento realizado detalhando a montagem dos ensaios de interação e a definição dos contextos de avaliação a partir destas três abordagens distintas.

4 DESCRIÇÃO DO EXPERIMENTO

Este capítulo descreve o experimento realizado para investigar qual a influência exercida pelos diferentes contextos de avaliação nos resultados da avaliação de usabilidade da interface com o usuário de uma aplicação de Internet móvel para telefone celular, baseada em três diferentes abordagens para a técnica de ensaios de interação.

Neste experimento 36 usuários, divididos em três grupos de doze pessoas, participaram da avaliação de usabilidade da interface com o usuário de um portal WAP. Dois grupos realizaram os ensaios de interação dentro do laboratório, sendo que o primeiro grupo utilizou um emulador no computador e o segundo grupo utilizou um telefone celular. O terceiro grupo utilizou também um telefone celular, mas realizou os ensaios de interação fora do laboratório, em um ambiente mais próximo ao contexto de uso deste tipo de equipamento.

A seguir serão descritas todas as etapas para a montagem e realização destes ensaios de interação segundo o Método de Medida da Performance (apresentado no item 2.2.2). Os critérios para comparação dos desempenhos das avaliações também estão aqui definidos.

4.1 Definição do produto a ser avaliado

Os ensaios de interação foram realizados com o objetivo de avaliar a usabilidade geral da interface com o usuário de um conhecido portal de Internet móvel. Este portal disponibiliza o acesso a informações e serviços variados como programação de lazer, notícias, gerenciamento de contas de *e-mail*, entre outros. O conteúdo

deste portal foi considerado, a partir de uma análise comparativa realizada pela autora, o mais completo e atualizado entre todos os portais WAP disponíveis atualmente no país.

Segundo dados fornecidos pelos desenvolvedores do portal (por e-mail à autora) em novembro de 2003 este portal contava com uma base de 27,4 milhões de usuários tendo uma audiência média de 25 mil usuários/dia, podendo ser acessado em 18 estados brasileiros, através de 11 operadoras de telefonia celular.

4.2 Definição do contexto de uso

A aplicação que foi avaliada pode ser acessada por qualquer usuário que possua um telefone celular desde que o mesmo seja compatível com a tecnologia WAP.

Todos esses serviços podem ser consultados a qualquer momento a partir de qualquer lugar que esteja dentro da área de cobertura da operadora de telefonia móvel como aeroportos, shoppings, hotéis, restaurantes, escritórios ou residências.

Não é necessário nenhum pré-requisito para acessar este serviço e qualquer pessoa que possua um telefone celular pode ser considerada um potencial cliente.

4.3 Definição dos diferentes contextos de avaliação

Foram definidas três abordagens distintas para a realização dos ensaios de interação a partir dos contextos de avaliação. Estes contextos diferem somente quanto a dois aspectos: o equipamento utilizado para a navegação no portal WAP (telefone celular ou emulador) e as condições da avaliação, que compreendem o local de realização dos ensaios e os equipamentos utilizados para registrar a

interação. Os demais componentes dos contextos de avaliação são exatamente os mesmos para as três abordagens como ilustra a figura 34.

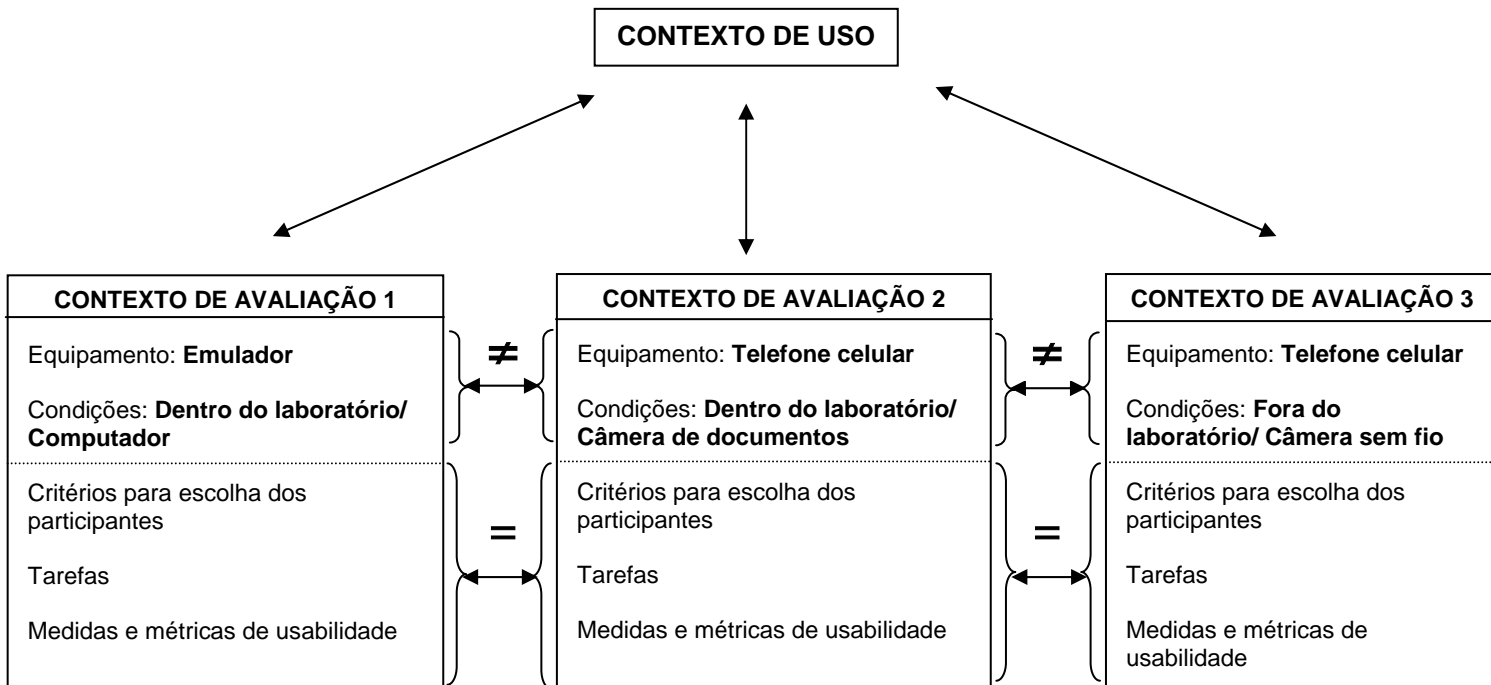


Figura 34: Componentes dos Contextos de Avaliação

4.3.1 Contexto de avaliação 1: dentro do laboratório utilizando um emulador

O primeiro grupo realizou os ensaios de interação dentro do laboratório utilizando o emulador no computador.

O computador conectado à Internet possuía monitor de 17" e mouse. Embora o emulador permitisse o uso do teclado o mesmo foi retirado para que o participante utilizasse apenas o mouse, desta forma a entrada de dados só foi possível clicando com o mouse nos botões do telefone celular que estão presentes na tela do computador, como em uma situação real de uso com o aparelho. Todo o sistema estava conectado, através de placa de vídeo e adaptador, a uma televisão de 20" integrada a um vídeo cassete. Desta forma, foi possível gravar em fita de vídeo toda

a interação do usuário com a interface. Um microfone também foi conectado para gravar o áudio da sessão. A figura 35 ilustra o ambiente deste ensaio de interação.



Figura 35: Ambiente do contexto de avaliação do emulador

Não havia câmeras filmando o participante, a imagem gravada em fita é somente a da tela do computador (figura 36).



Figura 36: Imagem da interação no contexto de avaliação do emulador

O avaliador se encontrava na mesma sala que o participante observando a interação através do monitor de televisão.

A partir de agora este contexto será referenciado como “contexto do emulador”.

4.3.2 Contexto de avaliação 2: dentro do laboratório utilizando o telefone celular e uma câmera de documentos

O segundo grupo realizou os ensaios de interação dentro do laboratório utilizando um telefone celular fixado em um tripé que estava posicionado sob uma câmera de documentos (modelo Canon RE-350).

O tripé foi desenvolvido especialmente para este trabalho com dimensões que permitissem o manuseio normal do aparelho como se o usuário o estivesse segurando nas mãos (figura 37). Os graus de liberdade da base de apoio também foram cuidadosamente definidos para possibilitar o melhor ângulo de visão tanto para o usuário quanto para a filmagem da tela e do teclado do telefone. O aparelho foi fixado no tripé utilizando fitas de velcro.



Figura 37: Tripé para fixação do telefone celular

Todo o sistema estava conectado através de placa de vídeo e adaptador do computador a uma televisão de 20" integrada a um vídeo cassete. Desta forma foi possível gravar em fita de vídeo toda a interação do usuário com a interface. Um

microfone também foi conectado para gravar o áudio da sessão. A figura 38 ilustra o ambiente deste ensaio de interação.



Figura 38: Ambiente do contexto de avaliação da câmera de documentos

Não havia câmeras filmando o participante, a imagem gravada em fita é somente a do telefone celular (figura 39).



Figura 39: Imagem da interação no contexto de avaliação da câmera de documentos

O usuário realizou o teste sentado e, em função do *design* do tripé, o telefone podia ser operado normalmente. Entretanto ele não poderia mover o tripé de sua posição inicial em relação à câmera de documentos e nem alterar a posição do telefone em relação ao tripé, pois o telefone celular não deveria sair da área de visualização da câmera.

O avaliador e o participante se encontravam na mesma sala. Toda a interação foi observada pelo avaliador através do monitor de televisão.

A partir de agora este contexto será referenciado como “contexto da câmera de documentos”.

4.3.3 Contexto de avaliação 3: fora do laboratório utilizando o telefone celular e uma minicâmera sem fio

O terceiro grupo realizou os ensaios de interação fora do laboratório em um local muito similar ao interior de um Shopping Center. Este local possui uma praça central com muitas mesas e cadeiras espalhadas e ao seu redor encontram-se várias lojas e restaurantes. Este foi o ambiente escolhido por três principais razões:

- É um ambiente que reproduz vários possíveis locais pertencentes ao contexto de uso deste tipo de equipamento como aeroportos, lojas, restaurantes, hotéis, etc.
- Apesar de não ser um ambiente totalmente fechado, é um local coberto e iluminado artificialmente, possibilitando a realização dos testes à noite independentemente das variações climáticas, o que não seria sempre possível se o ambiente fosse totalmente ao ar livre. A possibilidade de realização dos testes à noite é imprescindível uma vez que para uma grande maioria dos usuários

representativos do público alvo este era o único período disponível para a participação nos testes.

- É um ambiente seguro, requisito necessário não só visando o bem estar do participante e do avaliador mas também em função dos equipamentos que foram utilizados durante o teste.

Para filmar a interação com o equipamento em um ambiente externo é necessária a utilização de equipamentos como câmeras e microfones móveis e portáteis e que funcionem de forma autônoma, ou seja, que não precisem estar conectados à rede elétrica durante a realização dos testes.

Devido à dificuldade de encontrar no Brasil um sistema adequado e à impossibilidade de importá-lo em função do alto custo, foi necessário desenvolver uma solução especialmente para este trabalho que permitisse filmar a realização dos testes fora do laboratório.

O sistema desenvolvido possibilita a filmagem da tela e do teclado do telefone celular em qualquer ambiente, independentemente da necessidade de conexão à rede elétrica.

O sistema é composto de uma minicâmera sem fio com microfone embutido que se comunica via radiofrequência com um receptor, que pode estar localizado a uma distância de até 100 metros da câmera. Estes equipamentos operam com baterias de 9V e 12V respectivamente e permitem gravar a imagem colorida e também o áudio da sessão.

Foi necessário desenvolver também um sistema de fixação da câmera, do telefone celular e da bateria. Este sistema foi especialmente projetado para que o usuário pudesse operar o aparelho normalmente e ao mesmo tempo fosse possível gravar as imagens da interação com qualidade. A câmera foi fixada no suporte de

modo que a sua movimentação fosse solidária à movimentação do telefone. O suporte possui vários graus de liberdade o que permite a movimentação da câmera possibilitando vários ajustes diferentes com o objetivo de obter a imagem desejada do telefone sem prejudicar a visão do usuário e a qualidade da filmagem. O telefone e a bateria foram fixados no suporte com fitas de velcro (figura 40).

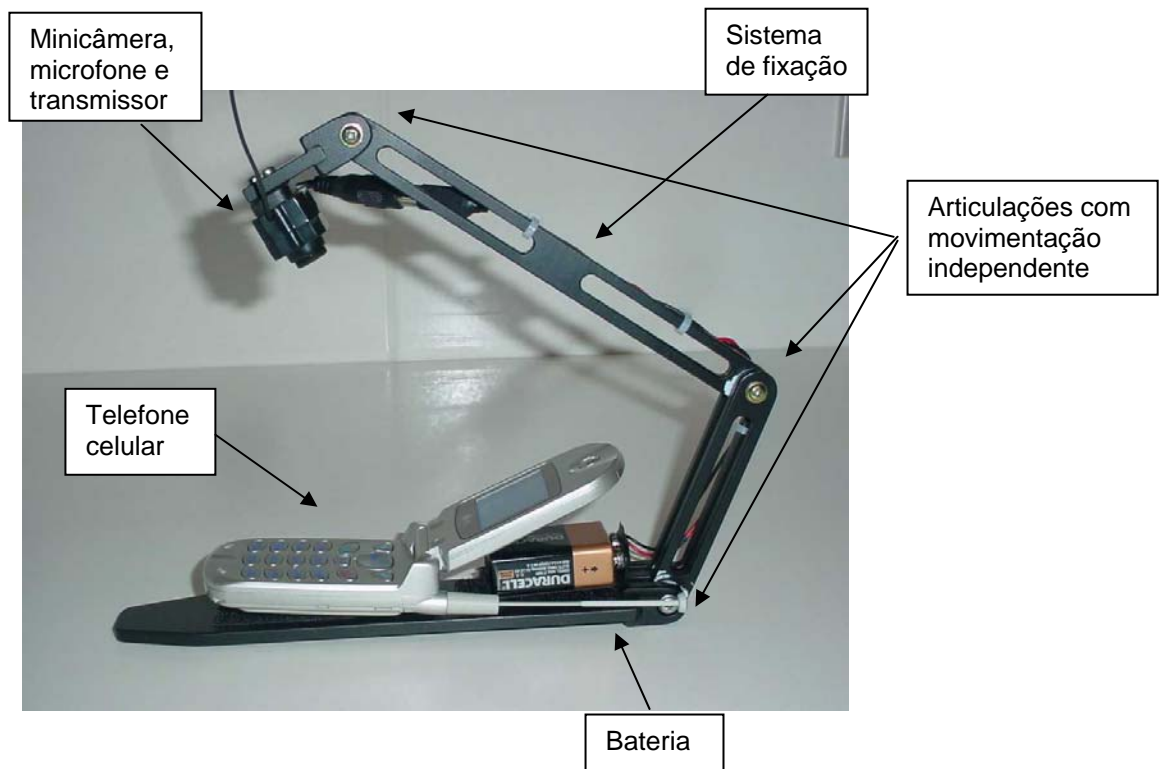


Figura 40: Sistema da minicâmera sem fio

A minicâmera envia os sinais de áudio e vídeo através de radiofreqüência a um receptor conectado a uma câmera filmadora digital portátil, que também opera com bateria e grava em fita as imagens da interação (foto 41).

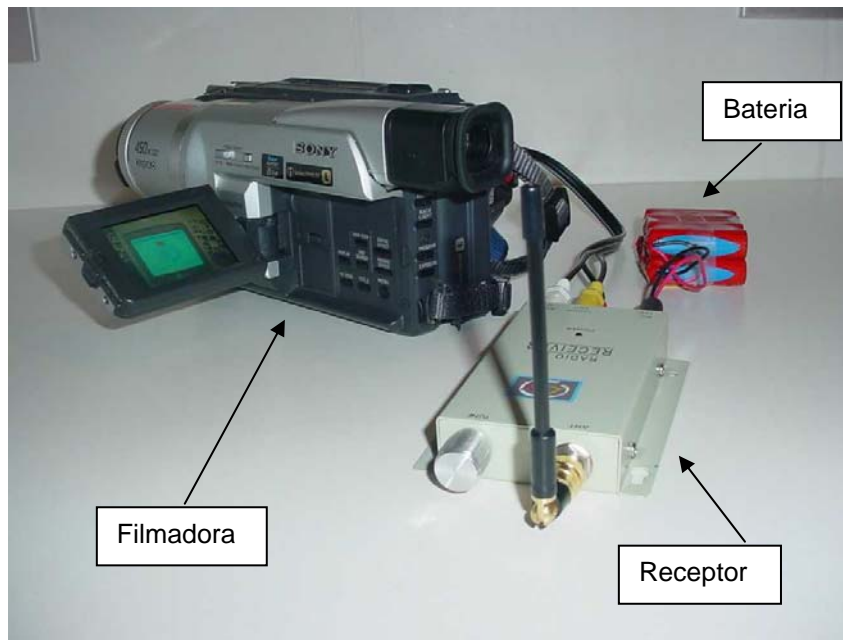


Figura 41: Sistema de recepção e gravação

O participante realizou o teste sentado, segurando o sistema formado pela minicâmera e o telefone nas mãos. Como, após ajustado, o telefone estava fixo e solidário em relação à minicâmera, o participante podia se movimentar e agir normalmente. Esta condição contribuiu para reproduzir com maior fidelidade o contexto real de uso sem prejudicar a filmagem. A figura 42 ilustra o ambiente deste ensaio de interação.



Figura 42: Ambiente do contexto de avaliação da minicâmera sem fio

O avaliador estava sentado próximo ao participante observando a interação através do display da câmera de vídeo. Não havia câmeras filmando o participante, a imagem gravada em fita é somente a do telefone celular (figura 43).

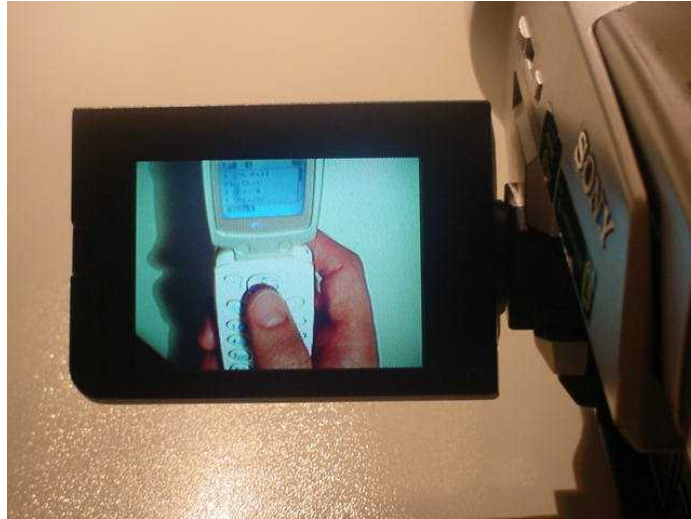


Figura 43: Imagem da interação no contexto de avaliação da minicâmera sem fio

Posteriormente, quando da análise das fitas no laboratório, a filmadora foi conectada à televisão permitindo que a imagem fosse reproduzida em uma tela maior facilitando a observação para o avaliador.

Este contexto será referenciado daqui para frente como “contexto da câmera sem fio”.

4.3.4 O emulador e o telefone celular

Vários modelos de telefones celulares e diversos emuladores foram pesquisados com o objetivo de encontrar o par que apresentasse o maior número possível de similaridades em termos de disposição e funcionalidade das teclas e da visualização da aplicação. Com base neste critério após uma extensa pesquisa

foram escolhidos um emulador da Openwave com uma imagem de telefone padrão e o software UP Simulator Openwave v.4.1.1 (Openwave, 2003) na versão compatível com a versão do *browser* do telefone celular selecionado.

O modelo de telefone celular escolhido foi o BD-4000 da LG que opera na área de cobertura da Vivo, usando a rede 2,5G na tecnologia CDMA 1X.

A disposição e a funcionalidade das teclas para entrada de dados, seleção e navegação são rigorosamente as mesmas no emulador e no telefone celular (figura 44).

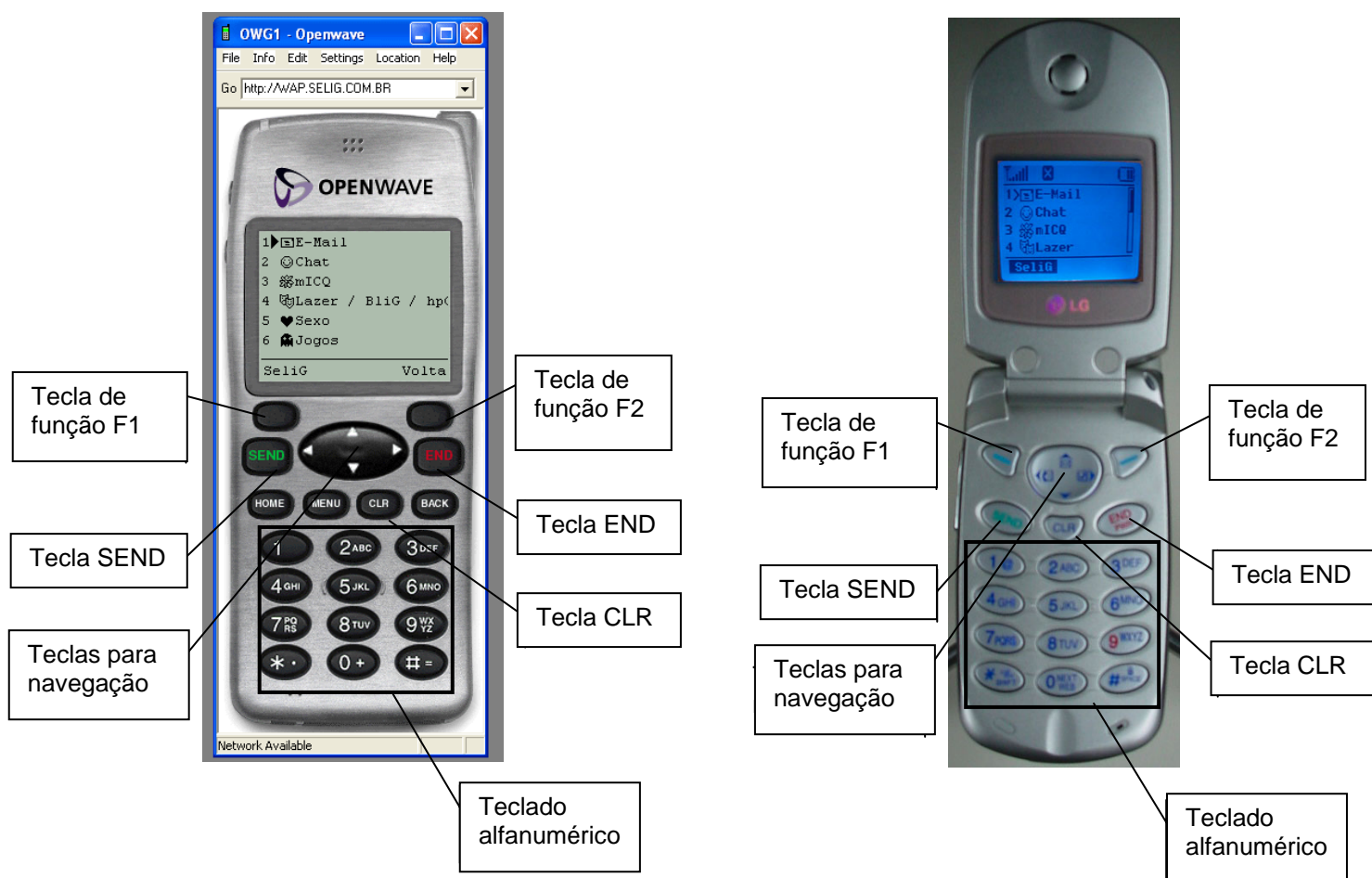


Figura 44: Emulador e telefone celular

Os textos e menus da aplicação são apresentados na tela seguindo a mesma disposição, entretanto há duas diferenças principais em relação às duas telas: o

número de linhas na tela e a barra de rolagem. O emulador apresenta na tela duas linhas a mais de texto em relação ao telefone celular e a indicação da barra de rolagem de texto que aparece no canto direito do telefone celular não existe no emulador (figura 45).

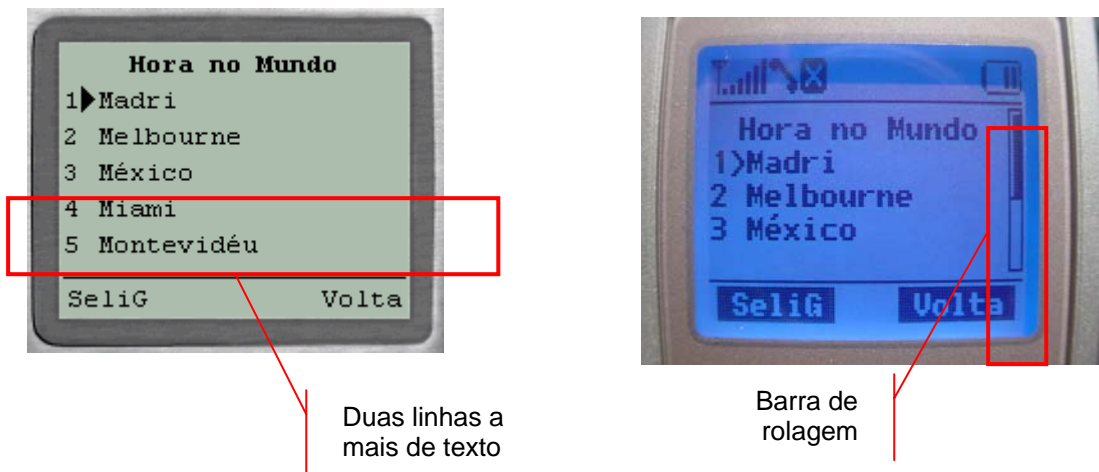


Figura 45: Diferenças entre as telas do emulador e do telefone celular

4.3.5 Definição da amostra de participantes

Foram selecionados ao todo 36 participantes, divididos em três grupos de 12 pessoas. Segundo Rubin (1994) e Spyridakis (1992) 10 pessoas é o número mínimo de participantes por grupo de teste se houver a necessidade de analisar os resultados estatisticamente.

Cada um dos usuários participou de um único ensaio de interação e, como recomendado em (Nielsen, 1993a), os participantes foram distribuídos aleatoriamente formando três grupos homogêneos.

A escolha dos usuários deve sempre ser realizada de forma que eles sejam tão representativos quanto possível em relação aos usuários reais do sistema (Rocha, 2003). Devido à inexistência de estudos que identifiquem o perfil do usuário WAP no

Brasil optou-se por seguir o perfil definido em estudos similares como (Nielsen, 2000). Os participantes deveriam atender aos seguintes critérios:

- Usuários de telefone celular (Nielsen, 2000; Weiss, 2000). Qualquer pessoa que possua um aparelho de telefone celular pode ser considerado um usuário em potencial dos serviços de dados.

- Não são usuários do serviço WAP como em (Nielsen, 2000; Chittaro, 2001; Vyas, 2002). A colocação de (Chittaro, 2001, p.6) se aplica à realidade brasileira atualmente, para o autor:

...a escolha de usuários novatos se justifica uma vez que o acesso a dados através do telefone celular ainda é uma tecnologia introduzida recentemente, sendo raros os usuários experientes e uma grande prioridade atrair novos usuários para esse tipo de serviço.

- Usuários com experiência no uso do computador. Um dos contextos de avaliação fará uso de um emulador no computador, portanto é importante que o participante já esteja familiarizado com o seu uso para que dificuldades com o mouse e o teclado não prejudiquem o resultado da avaliação. Para Nielsen (1993a) se o usuário não está treinado a usar o mouse e outros dispositivos necessários à interação estas dificuldades serão as principais enfrentadas pelo participante e impedirão que se obtenha qualquer informação sobre a usabilidade do diálogo com a interface.

- Não são usuários do modelo de telefone celular utilizado nos testes. Como recomendado por Nielsen (2000) isto evita que haja diferentes níveis de experiência em relação ao manuseio do aparelho.

- Usuários entre 21 e 40 anos de idade. Esta faixa etária foi a escolhida pois ela representa a maior concentração de usuários de telefone móvel no Brasil atualmente, correspondendo a 61% do total (ANATEL, 2003).

A tabela 1 apresenta a distribuição dos participantes de acordo com a faixa etária, o sexo e o contexto de avaliação.

Tabela 1: Distribuição dos participantes por faixa etária e sexo

Faixa etária	Contexto do emulador		Contexto da câmera de documentos		Contexto da câmera sem fio	
	Homens	Mulheres	Homens	Mulheres	Homens	Mulheres
21 a 30 anos	3	3	3	3	3	3
31 a 40 anos	3	3	3	3	3	3
Total	12		12		12	
Total geral	36					

- Os usuários deveriam ter diferentes formações e ocupações. Como constatado por Rocha (2003, p.204) “não é conveniente testar uma interface voltada para o público em geral e utilizar estudantes de computação como grupo de teste: eles certamente não são representativos da população de usuários alvo”, embora este seja um hábito comum entre pesquisadores como (Rao,2000; Brewster, 2002).

4.3.6 Definição do *script* das tarefas

A cada usuário foi solicitada a execução das mesmas sete tarefas. Estas tarefas foram selecionadas em função de dois critérios: (a) era necessário que fossem tarefas representativas do uso deste tipo de serviço de Internet móvel como encontrar informações sobre esporte, lazer e serviços de *e-mail* (Nielsen, 2000; Helyar, 2002; Weiss, 2001; Sears, 2003). (b) dentre estas tarefas foram escolhidas aquelas que poderiam identificar o maior número de problemas de usabilidade. Estes problemas já haviam sido apontados em um diagnóstico sobre a usabilidade geral do sistema realizado previamente pela autora.

As tarefas eram independentes umas das outras e estavam todas inseridas em diferentes cenários. A utilização de cenários é importante para contextualizar a tarefa permitindo ao usuário que ele vislumbre uma possível condição real de uso daquela aplicação e criando portanto uma motivação adicional para que a tarefa seja executada (Rubin, 1994).

O quadro 3 apresenta as tarefas utilizadas nos ensaios de interação, inseridas em seus respectivos cenários.

Quadro 3: Descrição das tarefas

TAREFA	DESCRIÇÃO
1	Você está em um restaurante jantando com amigos e o assunto é futebol. Um de seus amigos é torcedor do Internacional e gostaria muito de saber quando será o próximo jogo de seu time no campeonato brasileiro. Você se prontifica a encontrar esta informação. Seu objetivo: Encontre a data, o horário e o adversário do próximo jogo do Internacional no campeonato brasileiro de futebol.
2	Você está sentado na sala de embarque do aeroporto aguardando seu vôo para o Rio de Janeiro. Você estará livre esta noite e gostaria de aproveitar para ir ao teatro. Seu objetivo: Encontre o nome de todas as peças que estão em cartaz nos teatros do Rio de Janeiro.
3	Você está participando de uma importante reunião de negócios quando um colega seu solicita informações atualizadas sobre o valor do dólar. Seu objetivo: Encontre o valor do dólar comercial, o percentual de alta ou baixa e a data e hora a que essas informações se referem.
4	Você está com um cliente visitando uma empresa no bairro do Cabral. Você não conhece muito bem esse bairro e gostaria de saber quais são as opções de restaurante para que vocês possam almoçar/jantar agora. Seu objetivo: Encontre o nome, o endereço e os horários de funcionamento de um restaurante no bairro do Cabral.
5	Você percebe que há uma mensagem de voz em seu telefone celular. É de um amigo seu que está morando em Nova York. Ele pede que você retorne a ligação em um determinado número de telefone ainda hoje às 21:00h pontualmente que ele estará aguardando. Você não está muito certo quanto à diferença de horário entre Curitiba e Nova York e gostaria de saber a que horas você deve fazer essa ligação. Seu objetivo: Encontre o horário em Nova York agora.
6	Eu liguei para você ontem e disse que enviaria um e-mail com as informações sobre a nossa reunião. Você saiu de casa atrasado hoje e não teve tempo de abrir seus e-mails no computador. Agora você está na rua e lembra que precisa ler este e-mail porque a reunião pode ter sido marcada para hoje. Seu objetivo: Encontre o e-mail que eu enviei a você, leia-o inteiramente verificando qual o horário, a data e o local da nossa reunião. Seu e-mail é: wqoliveira@ig.com.br Sua senha de acesso é: manaus
7	Você leu o e-mail anterior e está de acordo com a data e o horário marcados. Você precisa agora confirmar a sua presença na reunião. Seu objetivo: Responda a este e-mail com o texto: Ok, confirmado!

4.3.7 Escolha das medidas e métricas de usabilidade

4.3.7.1 Usabilidade para os computadores de mão

A definição de usabilidade para os computadores de mão adotada neste trabalho foi a da ISO 9241-11 como sugerido em Hiltunen (2002). Segundo Jokela (2003) a ISO 9241-11 está se tornando a principal referência para a usabilidade pois, além de ser largamente reconhecida na literatura, foi a definição de usabilidade escolhida para a elaboração do CIF-Common Industry Format for Usability Test Reports (CIF, 2001), que fornece um modelo padrão para geração de relatórios de testes de usabilidade. Para Jokela (2003, p.53) “esta é uma indicação muito clara da importância desta definição uma vez que a criação do CIF foi apoiada por um grande número de empresas”.

Segundo Ericsson (2001) o usuário móvel é muito mais impaciente e exigente que o usuário do computador de mesa. Isto se dá em função do contexto de uso dos computadores de mão, normalmente ambientes dinâmicos nos quais a interação com o equipamento é uma entre as diversas tarefas que o usuário está realizando (Chincholle, 2002b). Neste contexto a usabilidade não é somente uma questão de facilidade de uso da interface, mas está relacionada a toda a extensão na qual o produto possa atender às necessidades do usuário lhe proporcionando funcionalidade, altos índices de desempenho, confiabilidade e facilidade de uso apropriados (Bevan, 1997). Este é exatamente o enfoque definido pela ISO 9241-11(1998) e operacionalizado pelo método MUSiC (Macleod, 1997; Bevan,1998).

A ISO 9241-11 (1998) define usabilidade como:

“a eficiência, a eficácia e a satisfação com as quais determinados usuários realizam determinadas tarefas em um determinado contexto de uso”.

Uma vez determinadas as componentes da usabilidade por esta definição, foi utilizado o Método de Medida da Performance (Cooper, 1995), parte integrante do projeto MUSiC (Macleod, 1997), para estabelecer as definições e os métodos de medidas dessas componentes.

As medidas de performance a partir das quais derivaram as métricas de usabilidade são obtidas a partir da análise do resultado da tarefa produzido pelo usuário e da fita de vídeo da sessão.

4.3.7.2 Eficácia da tarefa

Para cada uma das tarefas foram definidos critérios para medir a qualidade e a quantidade do resultado da tarefa em função do objetivo atingido pelo usuário. A eficácia foi então calculada pela fórmula:

$$\text{Eficácia} = \text{qualidade} \times \text{quantidade} / 100 \%$$

O quadro 4 apresenta os critérios de qualidade e quantidade definidos para cada uma das tarefas.

Quadro 4: Definição de qualidade e quantidade para as tarefas

TAREFA	OBJETIVO	QUALIDADE	QUANTIDADE
1	Encontrar a data, o horário e o adversário do próximo jogo do Internacional no campeonato brasileiro de futebol.	Encontrar o card que apresenta a lista dos próximos jogos.	Percentual da informação encontrada (35% data, 35% horário, 30% adversário).
2	Encontrar o nome de todas as peças que estão em cartaz nos teatros do Rio de Janeiro.	Encontrar o card que apresenta a lista das peças de teatro no Rio de Janeiro.	Percentual da informação encontrada em função do número de peças da lista.
3	Encontrar o valor do dólar comercial, o percentual de alta ou baixa e a data e hora a que essas informações se referem.	Encontrar o card que apresenta as informações sobre o dólar.	Percentual da informação encontrada (25% cada informação).
4	Encontrar o nome, o endereço e os horários de funcionamento de um restaurante no bairro do Cabral.	Encontrar o card que apresenta as informações sobre um restaurante no Cabral.	Percentual da informação encontrada (35% nome, 35% endereço, 30% horários de funcionamento).
5	Encontrar o horário em Nova York agora.	Encontrar o card que apresenta o horário para a cidade de Nova York.	100% se achou o horário correto.
6	Encontrar o e-mail que eu enviei a você, ler o e-mail inteiramente verificando qual o horário, a data e o local da nossa reunião.	Abrir o e-mail correto.	Percentual da informação encontrada (35% data, 35% horário, 30% local).
7	Responder ao e-mail lido com o texto: Ok, confirmado!	É considerada a qualidade máxima se o texto enviado correspondia ao texto solicitado. Para cada caracter digitado incorretamente no texto enviado a qualidade era reduzida em 5%.	100%, se enviou somente um e-mail para o endereço correto. Zero, se não enviou nenhum e-mail. Para cada e-mail adicional enviado ao mesmo endereço ou a outro endereço era descontado da quantidade um valor de 50%.

4.3.7.2.1 Tarefa completada

Completar ou não a tarefa está relacionado a verificar se os objetivos da tarefa foram atingidos (Cooper, 1995), portanto este valor está diretamente relacionado à eficácia da tarefa.

Desta forma, ficou estabelecido o critério de que se a eficácia da tarefa for maior ou igual a 50% será considerado que a tarefa foi completada, ou seja, considera-se que o usuário completou pelo menos metade dos objetivos propostos em termos de qualidade e quantidade.

Para cada tarefa é atribuído um valor que indica se o participante a completou (valor 1) ou não (valor 0).

$$\text{Tarefa completada} = \begin{cases} 1, & \text{se eficácia} \geq 50\% \\ 0, & \text{se eficácia} < 50\% \end{cases}$$

Ao final, a partir da soma destes valores, é possível medir o percentual de participantes que conseguiram completar os objetivos da tarefa. Este percentual pode ser calculado para todas as tarefas.

$$\text{Percentual de participantes que completaram a tarefa} = \frac{\sum (\text{tarefa completada}) \times 100\%}{12}$$

4.3.7.3 Eficiência do usuário

O tempo total de execução da tarefa foi medido e a partir dele foi calculada a eficiência através da fórmula:

$$\text{Eficiência} = \text{eficácia} / \text{tempo total da tarefa}$$

4.3.7.4 Eficiência relativa do usuário em relação ao usuário especialista

A eficiência do especialista foi considerada a média entre a eficiência de três usuários especialistas que executaram as mesmas tarefas, com os mesmos equipamentos e nos mesmos contextos que os demais participantes.

A eficiência relativa foi calculada pela fórmula:

$$\text{Eficiência relativa} = (\text{eficiência do usuário} / \text{eficiência do especialista}) \times 100\%$$

4.3.7.5 Período produtivo

O período produtivo foi definido em função do tempo considerado produtivo durante a execução da tarefa e foi calculado pela fórmula:

$$\text{Período produtivo} = \frac{(\text{tempo total} - \text{tempo ajuda} - \text{tempo busca} - \text{tempo snag})}{\text{tempo total}} \times 100\%$$

O período produtivo foi obtido subtraindo do tempo total de duração da tarefa o tempo perdido com as ações improdutivas. Durante a análise das fitas de vídeo a classificação das ações do usuário em relação ao tempo de execução da tarefa foi baseada no esquema representado na figura 46.

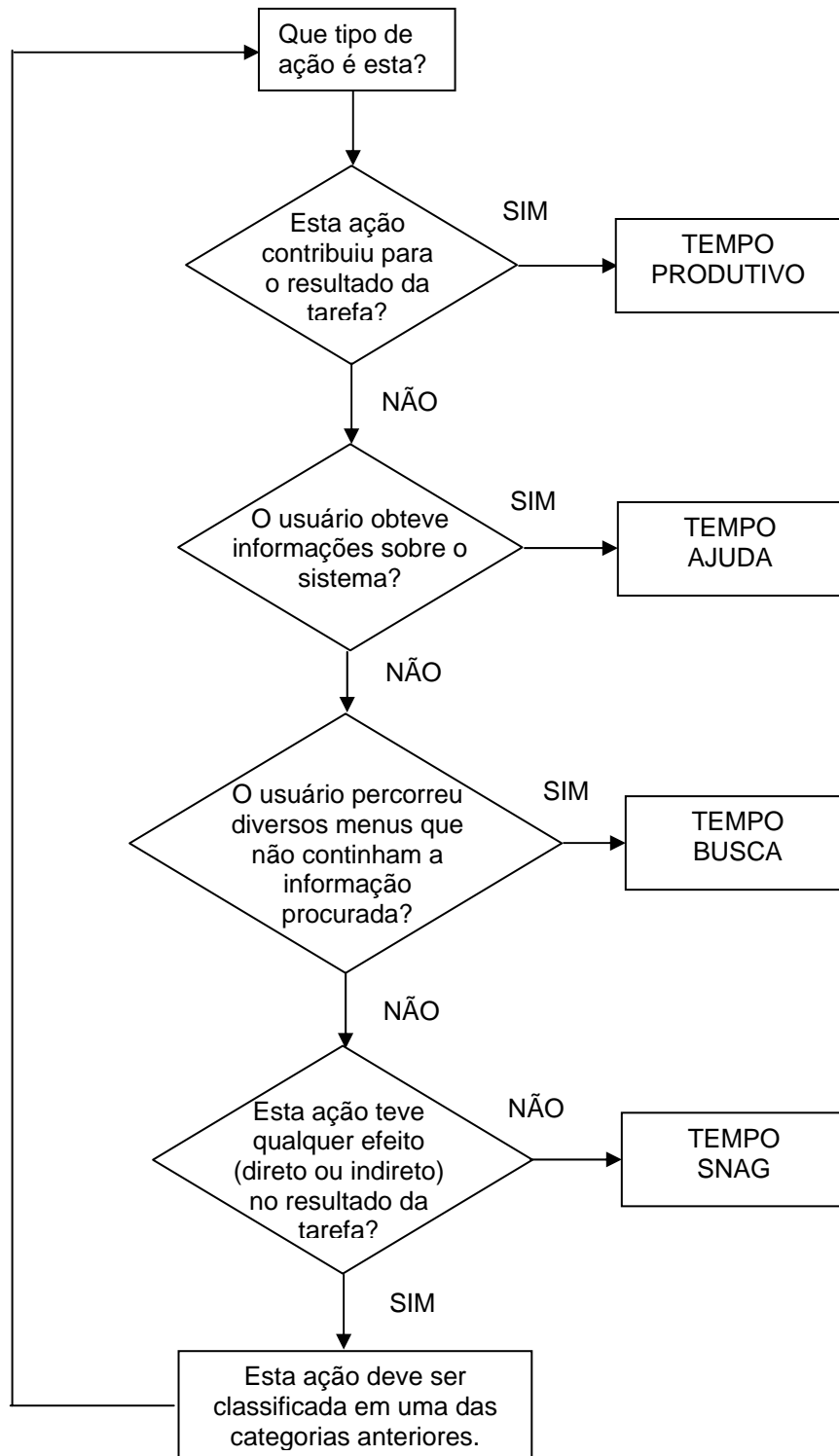


Figura 46: Árvore de decisão para categorizar as ações do usuário em relação ao tempo de execução da tarefa

Fonte: Adaptado de (Cooper, 1995)

Não foi definido um tempo limite para que o usuário completasse a tarefa. Ele poderia executar a tarefa até que ele a considerasse concluída como recomendado em (Cooper, 1995). Entretanto, alguns participantes em algumas tarefas, devido às dificuldades enfrentadas, desistiram de completá-las. A eficácia nestes casos foi considerada zero, pois eles não haviam atingido nenhum percentual dos objetivos.

Quando a eficácia é zero há uma regra no Método de Medida da Performance (Cooper, 1995) que estabelece que o período produtivo nestes casos também deve ser considerado zero. Esta regra baseia-se na definição de que período produtivo é o tempo despendido pelo usuário em direção ao cumprimento dos objetivos da tarefa. No caso da eficácia zero, como nenhum objetivo foi cumprido, não pode ser considerado nenhum período produtivo.

Nestas situações não há a necessidade de uma análise formal da fita de vídeo pois não são medidos os tempos, entretanto a eficácia zero indica que o usuário enfrentou muitos problemas e portanto a fita precisa ser cuidadosamente analisada para identificar os incidentes ocorridos.

4.3.7.6 Satisfação do usuário

As medidas subjetivas sobre a satisfação do usuário em relação ao uso do sistema foram obtidas através do questionário SUS-System Usability Scale (Brooke, 1996; SUS, 2003). O SUS é um questionário fechado composto de dez questões, usando a escala Likert, que cobrem diversos aspectos da usabilidade do sistema como complexidade, necessidade de treinamento e consistência. O resultado da avaliação do SUS é um valor numérico entre zero e cem.

Além de poder ser respondido rapidamente ele é um questionário geral que pode ser aplicado a um grande número de sistemas e pôde ser facilmente adaptado à aplicação WAP que estava sendo avaliada (anexo A). O SUS está disponível na Internet (SUS, 2003) e pode ser utilizado livremente desde que seja mencionada a fonte. Os questionários (SUMI, 2003) e (QUIS, 2003) são questionários mais longos e complexos, além disto eles não podem ser utilizados livremente, precisam ser licenciados ou adquiridos junto aos desenvolvedores.

4.3.7.7 Avaliação da carga de trabalho

Além das diferenças entre os equipamentos utilizados para realizar a navegação (emulador X telefone celular) e para registrá-la (computador X câmera de documentos X câmera sem fio), os usuários estavam expostos a diferentes ambientes (dentro do laboratório X fora do laboratório) e situações (local silencioso X local de movimento e barulho). Desta forma, decidiu-se também avaliar qual a carga de trabalho imposta ao participante pelo contexto de avaliação como em (Brewster, 2002).

A carga de trabalho pode ser definida como um conceito hipotético que representa o custo para o ser humano em conseguir um determinado nível de rendimento (Arquer, 2001). A experiência subjetiva da carga de trabalho resume as influências de diversos fatores além das demandas objetivas impostas pela tarefa. Assim a carga não é uma característica inerente à tarefa, mas é o resultado da interação entre os requisitos da tarefa, as circunstâncias em que ela se realiza, as capacidades, a conduta e a percepção do participante. Esta carga de trabalho pode

ser medida e os métodos subjetivos de avaliação de carga de trabalho são de ampla aplicação devido a sua facilidade de uso, validade e aceitação (Valdehita, 1999).

Com o objetivo de avaliar a carga de trabalho subjetiva dos participantes em relação ao contexto de avaliação, no final do ensaio de interação foi aplicado o método NASA Task Load Index (NASA TLX, 1986), método recomendado pela ISO 9241-11 (1998) e pelo MUSiC (Macleod, 1997).

O NASA-TLX é um procedimento de valoração multidimensional que fornece uma pontuação global para a carga de trabalho. Este valor é obtido a partir da média ponderada das pontuações em seis fatores: demanda mental, demanda física, demanda temporal, esforço, performance e frustração. A aplicação deste método se dá em duas etapas: pontuação e ponderação. Na fase de pontuação o participante deve estimar a carga de trabalho em relação a cada um dos seis fatores marcando uma posição em uma escala de 0 a 100.

Para cada fator é definida uma escala dividida em intervalos de 5 unidades, e limitada nos extremos por descritores como alto/baixo. Na fase de ponderação o participante define os fatores de carga. São apresentados aos participantes quinze comparações binárias dos seis fatores para os quais ele deve escolher, entre cada par, qual ele percebe como maior fonte de carga. Desta forma se obtém um peso para cada um dos fatores em função do número de vezes que ele apareceu. Cada fator pode receber peso zero (não apareceu nenhuma vez e portanto não é considerado relevante) até peso cinco (o fator foi escolhido todas as vezes e, portanto, é considerado a fonte de carga mais importante). Desta forma, além do valor geral da carga de trabalho, é possível identificar quais são os fatores determinantes para a geração desta carga.

Assim, a partir dos valores obtidos com a aplicação do NASA-TLX, é possível avaliar os níveis de carga de trabalho percebidos pelos participantes em cada um dos contextos de avaliação tanto em valores gerais quanto por fatores geradores de carga.

4.4 Definição dos critérios para comparação dos desempenhos das avaliações

Além de comparar os contextos de avaliação a partir das medidas de usabilidade e da carga de trabalho percebida pelo usuário como descrito anteriormente, os três contextos de avaliação serão comparados também em função do número de problemas de usabilidade identificados, da gravidade desses problemas e do número de horas gastas na realização das avaliações. A descrição de todos estes critérios se encontra a seguir.

4.4.1 Atribuição de graus de severidade aos problemas de usabilidade

Lavery (1997, p.254) define um problema de usabilidade como:

Um problema de usabilidade é um aspecto do sistema ou uma demanda imposta ao usuário que torna desagradável, ineficiente, oneroso ou impossível para o usuário atingir os seus objetivos em uma situação normal de uso.

Esta definição explicita claramente os dois aspectos de um problema de usabilidade: (a) a causa, referente à interface, ou seja, um aspecto do sistema e (b) a dificuldade dela resultante, referente ao usuário, ou seja, um incidente na interação.

Quando a avaliação envolve a observação de usuários, o que o avaliador observa são os problemas que ocorrem durante a interação, são estes incidentes na

interação que vão permitir ao avaliador identificar posteriormente os aspectos da interface que o causaram. Segundo Rubin (1994, p.276) “este é um ponto de transição da orientação à tarefa para a orientação ao produto”. É importante salientar que esta não é uma relação de um para um, mas sim de muitos para muitos, pois um único incidente na interação pode ter sido causado por vários aspectos da interface e um único aspecto da interface pode ser a causa de vários incidentes na interação.

A severidade dos problemas de usabilidade é um fator importante quando se comparam métodos de avaliação, pois ser capaz de encontrar um grande número de problemas não prioritários não torna um método melhor que outro (Desurvire, 1994). Portanto, para poder identificar a capacidade do método em identificar problemas mais graves, a cada problema de usabilidade identificado foi atribuído um grau de severidade definido em função de dois fatores: o impacto que o problema causou no usuário e a frequência com que ele ocorreu (Jeffries, 1991):

$$\text{Grau de severidade} = f(\text{impacto, número de usuários})$$

O impacto foi definido como uma escala de três pontos (Desurvire, 1994):

- Impacto 1 – o usuário fica um pouco confuso ou aborrecido e hesita rapidamente.
- Impacto 2 – o usuário é levado a um erro do qual ele consegue se recuperar.
- Impacto 3 – o usuário é levado a um erro do qual ele não consegue se recuperar, impossibilitando-o de concluir a tarefa com sucesso.

A frequência foi definida em função do número de usuários que enfrentaram aquele problema.

4.4.2 Definição das medidas e métricas de desempenho das avaliações

A qualidade das avaliações realizadas nos três diferentes contextos foi especificada em função das seguintes medidas e métricas:

- O número total de problemas de usabilidade identificados (Lavery, 1997).
- A soma dos graus de severidade de todos os problemas de usabilidade identificados (Jeffries, 1991). A partir desta métrica é possível identificar a capacidade efetiva de cada contexto de avaliação em detectar os incidentes de interação relacionados aos graus de impacto.
- A relação benefício/custo definida como a soma dos graus de severidade dividida pelo número total de horas gastas na avaliação (Jeffries, 1991).

4.5 Condução dos ensaios de interação

As avaliações foram distribuídas alternadamente entre os contextos e realizadas ao longo de duas semanas, sendo que a autora deste trabalho foi a única avaliadora responsável pela condução de todos os ensaios de interação.

Cada um dos ensaios de interação durou em média uma hora e meia. Todos os testes seguiram rigorosamente o mesmo roteiro que foi elaborado a partir de (Rubin, 1994; Dumas, 1999; Nielsen, 2003b). Este roteiro está descrito a seguir:

4.5.1 Introdução do participante ao ambiente do teste

- O participante foi recebido e lhe foi explicado como seria conduzida a avaliação, o que ele teria que fazer e quanto tempo duraria a sessão. É comum que

o participante chegue um pouco apreensivo e inseguro pois, apesar do primeiro contato quando do recrutamento, ele chega ao local do teste com muitas dúvidas. É muito importante que o avaliador forneça ao participante todas as informações sobre a condução da avaliação para que este possa se sentir à vontade. Na literatura (Nielsen, 1993a; Dumas, 1999; Rubin, 1994) é fortemente recomendado que o avaliador explique claramente ao participante que não é ele quem está sendo avaliado, e sim que o objetivo da avaliação é a aplicação. Durante a realização destes ensaios de interação em nenhum momento foi mencionada ao usuário a palavra “teste”. No recrutamento os participantes foram convidados a participar de uma “pesquisa” de usabilidade e não de um “teste” de usabilidade.

- Foram mostrados ao participante todos os equipamentos utilizados no teste, e também foi explicado que seu rosto não estaria sendo filmado e que todo o procedimento seria anônimo.
- Foi pedido ao participante que preenchesse um rápido questionário com alguns dados sobre o seu perfil. Um modelo de questionário para registrar o perfil do usuário pode ser encontrado no apêndice B.

4.5.2 Treinamento do participante

- O participante foi instruído quanto às principais funcionalidades do telefone celular e do emulador.
- Foram demonstrados os comandos básicos para iniciar a navegação no WAP.
- O participante foi introduzido ao NASA TLX. O método foi explicado e o participante analisou as folhas deste método que seriam preenchidas ao final do experimento.

4.5.3 Execução das tarefas

- As tarefas foram entregues ao participante, uma de cada vez, em ordem crescente de dificuldade. As tarefas foram lidas e em seguida foi entregue ao participante uma cópia em papel que ele poderia consultar sempre que julgasse necessário.

- Foi perguntado ao participante se ele tinha alguma dúvida e foi então dado início à execução da tarefa. Toda tarefa era iniciada na tela do menu principal do portal e o participante avisava quando considerava a tarefa como concluída. Ao concluir a tarefa o participante deveria ler em voz alta as informações encontradas.

- Toda a interação foi observada através do monitor de televisão (dentro do laboratório) e do visor da câmera filmadora (fora do laboratório) sem interferir na execução da tarefa. O avaliador anotou todas as observações feitas, bem como possíveis comentários e reações do participante. Ao final de cada tarefa foi realizada uma rápida entrevista com o participante para esclarecer possíveis dúvidas quanto aos incidentes de interação observados. Um modelo de instrumento para acompanhar a avaliação está reproduzido no apêndice C.

4.5.4 Questionários

- Foi solicitado ao participante o preenchimento do questionário de satisfação SUS para avaliar a satisfação do usuário.

- Foi solicitado ao participante o preenchimento do NASA TLX para avaliar a carga de trabalho.

- Foi solicitado ao participante o preenchimento do questionário 'Avaliação Geral sobre WAP' para avaliar a percepção do usuário em relação ao serviço WAP. Este questionário foi desenvolvido a partir do modelo apresentado por Nielsen (2000) e está reproduzido no apêndice D.

4.5.5 Debriefing

Ao final da avaliação foi realizada uma discussão com o participante para obter comentários gerais sobre a avaliação e esclarecimentos sobre as possíveis dificuldades que ele poderia ter enfrentado durante a interação.

Após o *debriefing* a avaliação estava encerrada.

4.6 Levantamento dos dados

Após a realização de todos os ensaios de interação foram quantificados os dados dos questionários aplicados (Perfil do Usuário, SUS, NASA-TLX, Avaliação Geral WAP) e cada uma das fitas de vídeo foi detalhadamente analisada.

4.7 Análise das fitas de vídeo

As sessões geraram um total de 32 horas de fita de vídeo que foram analisadas com os seguintes objetivos:

- Medir o tempo total de duração da tarefa.
- Medir o tempo gasto em ações improdutivas classificando e medindo separadamente o tempo de help, busca e snag.

- Analisar o resultado da tarefa executada pelo usuário para determinar com que eficácia os objetivos foram atingidos quantificando Qualidade e Quantidade de acordo com a definição destes critérios para cada uma das tarefas.
- Identificar os problemas de interação enfrentados pelo usuário classificando-os por grau de impacto. Nesta etapa eram utilizadas também as anotações feitas pelo avaliador durante os ensaios de interação com o usuário.

Para cada uma das tarefas a fita foi assistida pelo menos três vezes. Na primeira vez foi medido o tempo total que o usuário levou para executar a tarefa e analisado o resultado obtido para verificar se os objetivos foram ou não atingidos. Na segunda vez foi medido o tempo improdutivo, o que pode ser considerado uma das partes mais trabalhosas e demoradas uma vez que não pode ser feito de forma linear mas sim utilizando os controles de pausa, avanço e retrocesso da fita para marcar com precisão o início e o final de cada uma das categorias de tempo de ajuda, busca, e *snag*. A tarefa foi assistida ainda uma terceira vez para analisar os problemas de interação que foram identificados.

Os comentários mais relevantes dos usuários no *debriefing* após a realização da tarefa e na entrevista final também foram transcritos.

Um modelo de instrumento para auxiliar nesta análise das fitas está reproduzido no apêndice E.

Após a conclusão da análise das fitas todos os dados levantados foram colocados em tabelas que continham as fórmulas para calcular os valores da eficácia, eficiência, eficiência relativa e produtividade. O resultado foi um total de 21 tabelas (1 tabela para cada uma das 7 tarefas, para cada um dos 7 contextos de avaliação).

4.8 Análise dos problemas de usabilidade

Todos os problemas de usabilidade identificados foram analisados e classificados em fichas onde constava o nome e o número da tarefa, uma descrição e uma imagem da tela da localização do problema, a descrição do problema na interface (causa, identificada na interface pelo avaliador a partir dos incidentes na interação observados), a descrição do problema na interação (incidentes) com os respectivos impactos e número de usuários correspondes a cada incidente.

A partir destas fichas foi gerada uma lista com todos os problemas de usabilidade identificados juntamente com os seus respectivos graus de severidade.

4.9 Análise estatística

Para comparar os resultados das três situações experimentais foi empregado o método da Análise de Variância (ANOVA) (Ara, 2003).

Este método pode ser utilizado quando se deseja testar a igualdade de duas ou mais médias populacionais, baseado na análise das variâncias amostrais (Winter, 2002).

O método aqui utilizado é chamado de análise da variância de um critério uma vez que os grupos foram categorizados em função de uma única característica, ou critério, que neste estudo corresponde ao contexto de avaliação.

Ao comparar as médias dos resultados obtidos em cada um dos contextos de avaliação estaremos testando duas hipóteses:

H₀: Não existe diferença entre as médias (hipótese nula)

H₁: pelo menos uma das médias é diferente das demais (hipótese alternativa)

A análise de variância vai separar a variabilidade total nos dados em uma parte devida às diferenças entre as médias e outra devida às diferenças entre as observações no mesmo contexto.

O teste F (Winter, 2002) é utilizado para verificar se a variabilidade dos dados é decorrente do fato de que os mesmos pertencem a diferentes contextos de avaliação ou se é devido ao acaso (variabilidade residual).

A partir dessa análise será possível determinar se a hipótese nula deve ou não ser rejeitada. Se a hipótese nula não for rejeitada, conclui-se que não existem diferenças significativas entre as médias. Se a hipótese nula for rejeitada, isto indica que as médias não são iguais, entretanto não é possível determinar quais são as médias estatisticamente diferentes das demais. Nestes casos foi aplicado o teste de Tukey (Winter, 2002) para determinar quais são as médias diferentes entre si.

O resultado da análise estatística será apresentado como no exemplo a seguir:

$$"(\alpha = 5\%, F(x)=y, p=z)"$$

Estes valores devem ser interpretados da seguinte maneira:

- α é o nível de significância e mede a probabilidade de rejeitar H_0 quando ela é verdadeira.
- x é o valor tabelado da Distribuição F (para $\alpha = 5\%$) e depende do nível de significância do teste e dos graus de liberdade de tratamentos e de resíduos.
- y é o valor calculado de F para o teste em questão.

- z é a probabilidade de se obter um valor da estatística de teste no mínimo tão extremo quanto o que resultou dos dados amostrais, na suposição de a hipótese nula ser verdadeira. Ou seja, supondo que as médias sejam iguais, o valor de p indica qual a probabilidade de ocorrer ao acaso na amostra dos dados um valor muito extremo. Assim:

Quanto maior o p -valor, maior é a probabilidade de que os valores extremos tenham ocorrido ao acaso, então maior é a evidência de que não existem diferenças entre as médias, e neste caso diminui a certeza de rejeitar a hipótese nula.

Quanto menor o p -valor, menor é a probabilidade de que os valores extremos tenham ocorrido ao acaso, então maior é a evidência de que existem diferenças entre as médias, e neste caso a hipótese nula é rejeitada com maior certeza.

Podem ser considerados os seguintes valores:

- $p\text{-valor} < 0,01$ = elevada significância estatística, é uma evidência muito forte contra a hipótese nula.

- $0,01 < p\text{-valor} < 0,05$ = estatisticamente significativa, é uma evidência moderada contra a hipótese nula.

- $0,05 < p\text{-valor} < 0,1$ = estatisticamente significativa, é uma evidência fraca contra a hipótese nula.

- $p\text{-valor} > 0,1$ = evidência insuficiente contra a hipótese nula, ela não deve ser rejeitada.

5 RESULTADOS

5.1 Usabilidade medida nos três contextos de avaliação

5.1.1 Eficácia da tarefa

Os resultados obtidos para a eficácia das tarefas (figura 47) mostram que os usuários conseguiram completar 100% dos seus objetivos nas tarefas 2, 3, 4 e 6 nos três contextos de avaliação.

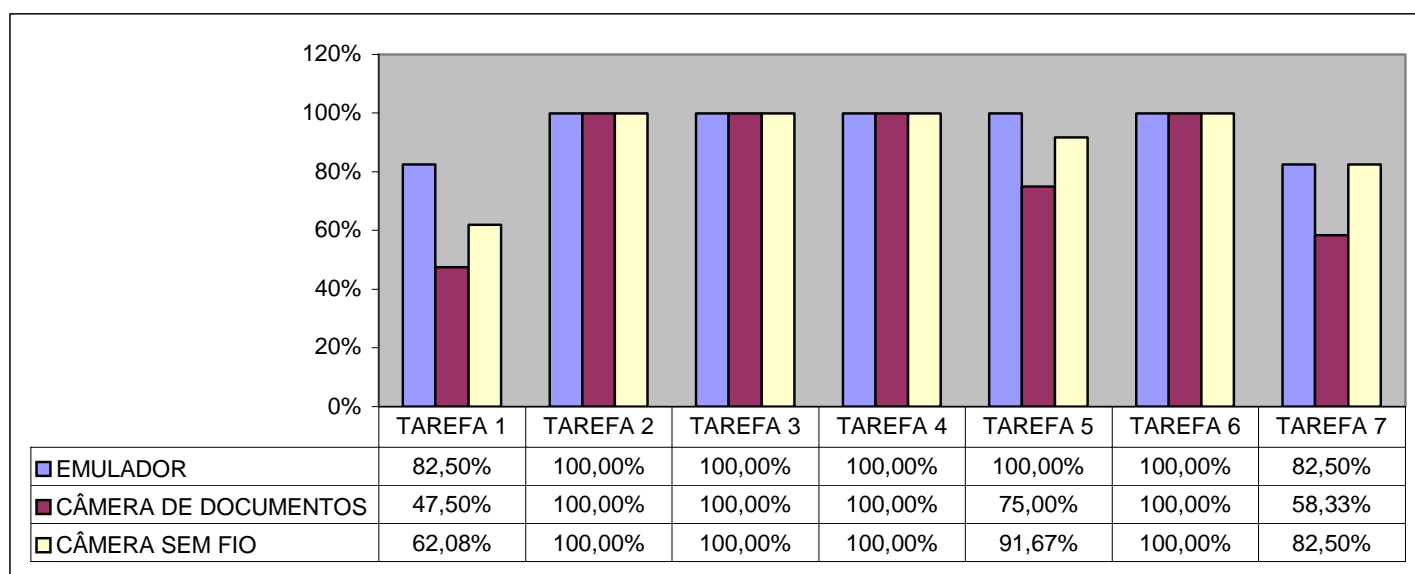


Figura 47: Resultados para a eficácia da tarefa

A tarefa1 apresentou uma eficácia estatisticamente menor no contexto da câmera de documentos em relação ao contexto do emulador ($\alpha=5\%$, $F(3,32) = 5,28$, $p=0,01$). Entre o contexto da câmera sem fio em relação aos contextos da câmera de documentos e do emulador não houve diferença estatística entre as médias para

esta tarefa. A figura 48 mostra que o percentual de usuários que completaram a tarefa 1 foi de 100% no contexto do emulador, caindo para 50% no contexto da câmera sem fio e para 41,67% no contexto da câmera de documentos.

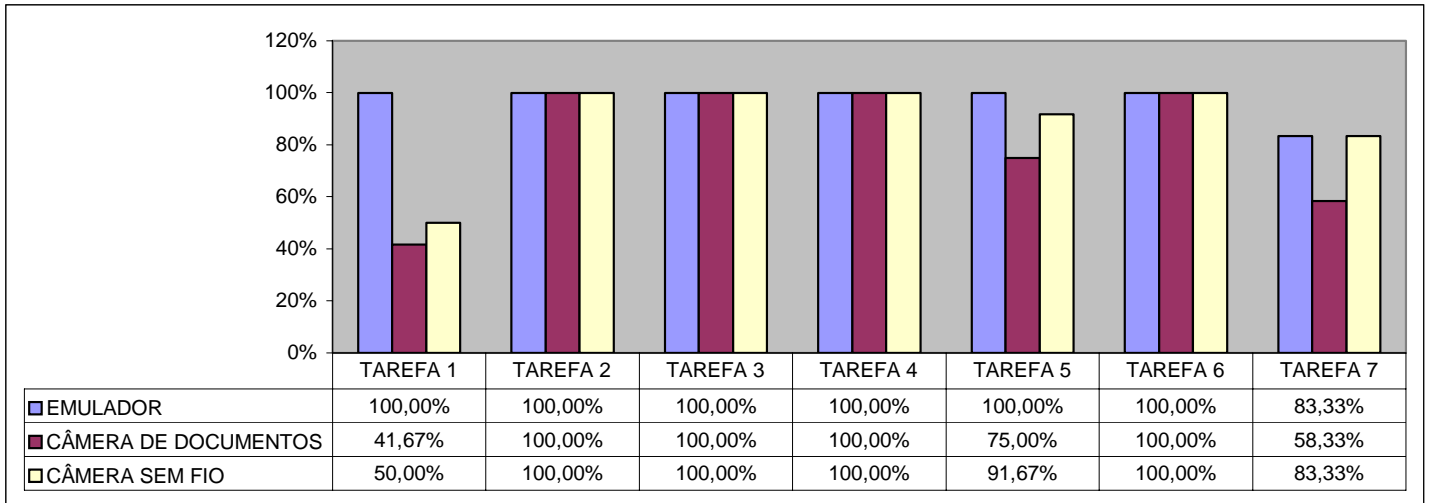


Figura 48: Percentual de usuários que completaram a tarefa

As diferenças entre as médias da eficácia nos três contextos de avaliação para a tarefa 5 não foram estatisticamente significantes ($\alpha=5\%$, $F(3,32)= 2,02$, $p=0,14$). No contexto do emulador 100% dos participantes completaram esta tarefa. Entretanto nos outros dois contextos alguns participantes desistiram da tarefa não chegando a completá-la, sendo que 25% dos participantes não completaram esta tarefa no contexto da câmera de documentos e 8% não completaram esta tarefa no contexto da câmera sem fio.

Para a tarefa 7 as diferenças entre as médias da eficácia para os três contextos de avaliação também não foram estatisticamente significantes ($\alpha=5\%$, $F(3,32)= 1,24$, $p=0,30$). Nos contextos do emulador e da câmera sem fio 16% dos participantes não conseguiram completar esta tarefa, para o contexto da câmera de documentos esse número subiu para 42% dos participantes.

Conforme os resultados apresentados, a análise estatística mostrou que não houve diferenças para a eficácia medida nos três contextos de avaliação com exceção da tarefa 1 que apresentou diferença entre as médias dos contextos do emulador e da câmera de documentos.

Para compreender melhor estes resultados é preciso analisar novamente a tarefa 1. Esta tarefa tinha como objetivos encontrar o adversário, a data e o horário do próximo jogo do Internacional no Campeonato Brasileiro de Futebol. Como todos os participantes conseguiram encontrar a tela que continha essas informações a qualidade para todos foi 100% e portanto o valor da eficácia foi determinado em função da quantidade da informação encontrada. A quantidade era calculada como 30% para o adversário correto, 35% para a data correta e 35% para o horário correto. A figura 49 ilustra dois exemplos de telas no emulador e no telefone celular para esta tarefa.



Figura 49: Telas da tarefa 1 no emulador e no telefone celular

Todos os participantes acertaram o adversário (Guarani) entretanto a maioria fez confusão com a data. Ao ler a informação os participantes acharam que a data seria dia 21/09 quando na realidade a data correta era 20/09, que não está mais visível na

tela. A mesma confusão aconteceu com o horário quando muitos pensaram que seria 16:00h e, na realidade, o horário correto era 18:00h.

No contexto do emulador 50% dos participantes encontraram todas as informações corretas, enquanto que os outros 50% acertaram o adversário e o horário, mas falaram a data errada. No contexto da câmera de documentos somente 8,3% (1 participante) conseguiu encontrar todas as informações corretas, 33,3% acertaram o adversário e o horário enquanto 58,3% acertaram somente o adversário. No contexto da câmera sem fio 41,6% encontraram todas as informações corretas, 8,3% acertaram o adversário e o horário enquanto 50% acertaram somente o adversário.

Essa diferença na eficácia para a tarefa 1 nos contextos do emulador e da câmera de documentos poderia ser atribuída à pequena diferença que existe entre o tamanho da tela do emulador e do telefone celular. Na tela menor do telefone o participante poderia ter uma dificuldade maior em identificar a data e o horário corretos. Entretanto esta diferença entre os tamanhos das telas não provocou uma diferença significativa entre as médias nos contextos do emulador e da câmera sem fio. Entre os contextos da câmera sem fio e da câmera de documentos, onde os participantes visualizavam a mesma tela do telefone celular, também não houve diferença significativa entre as médias. É possível concluir então que o que pode ter provocado a diferença para a eficácia entre os contextos do emulador e da câmera de documentos foi a soma de dois fatores: a tela menor do telefone celular e as condições de teste da câmera de documentos. Ambos os fatores tiveram um impacto negativo nos participantes que utilizaram a câmera de documentos em relação aos participantes que realizaram a avaliação no contexto do emulador.

Apesar de não haver diferenças estatisticamente significativas entre as médias da eficácia para as tarefas 5 e 7, o percentual de usuários que não conseguiram completar estas tarefas no contexto da câmera de documentos é maior em relação aos outros dois contextos de avaliação. Todos os usuários que não conseguiram completar estas tarefas foram usuários que, após um certo tempo enfrentando dificuldades, desistiram de continuar tentando e abandonaram a tarefa. Este índice de usuários que abandonaram as tarefas foi maior no contexto da câmera de documentos em relação aos dois outros contextos. Para os contextos da câmera sem fio e do emulador esses índices são os mesmos para a tarefa 7 e diferem somente em 8,3 % (1 usuário) para a tarefa 5. Estes resultados indicam que as condições da avaliação na câmera de documentos foram piores para o usuário fazendo com que eles desistissem de completar as tarefas com maior frequência em relação aos outros dois contextos de avaliação.

5.1.2 Eficiência do usuário

As médias da eficiência do usuário (figura 50) para as tarefas 1, 4 e 7 foram estatisticamente iguais nos três contextos de avaliação ($\alpha=5\%$, $F(3,32)=1,58$, $p=0,22$; $\alpha=5\%$, $F(3,32)=1,33$, $p=0,27$; $\alpha=5\%$, $F(3,32)=3,28$, $p=0,04$ respectivamente). A eficiência do usuário foi menor no contexto da câmera de documentos em relação ao contexto do emulador para as tarefas 2,3,5 e 6 ($\alpha=5\%$, $F(3,32)=9,91$, $p=0,0004$; $F(3,32)=9,01$, $p=0,0007$; $F(3,32)=4,27$, $p=0,022$; $F(3,32)=3,62$, $p=0,0378$ respectivamente).

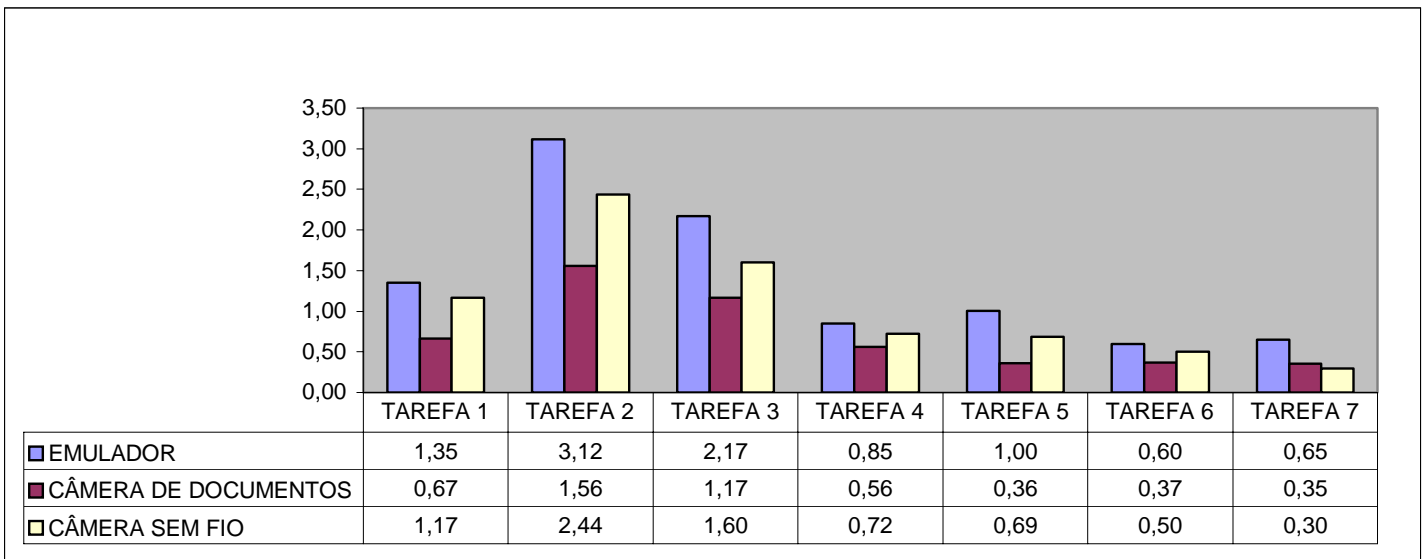


Figura 50: Resultados para a eficiência do usuário

Como a eficácia para as tarefas 2, 3 e 6 foi 100% nos três contextos, podemos concluir que a menor eficiência nessas tarefas decorre do fato de que no contexto da câmera de documentos os participantes levaram mais tempo para atingir os objetivos em relação aos participantes que realizaram os testes no emulador. Esta constatação é comprovada quando fazemos a ANOVA para os tempos médios nos três contextos e verificamos que o tempo no contexto da câmera de documentos é sempre maior para as tarefas 2,3 e 6 ($\alpha=5\%$, $F(3,32)=5,96$, $p=0,0061$; $F(3,32)=3,87$, $p=0,03$; $F(3,32)=5,11$, $p=0,01$). A menor eficiência para a tarefa 5 pode ser atribuída ao fato de que 3 participantes desistiram de completar esta tarefa no contexto da câmera de documentos e os que conseguiram completá-la levaram mais tempo neste contexto do que no contexto do emulador.

Entre os contextos da câmera de documentos e da câmera sem fio não houve diferença significativa para a eficiência das tarefas exceto para a tarefa 2 onde a eficiência do usuário foi menor também no contexto da câmera de documentos. Como a eficácia desta tarefa foi 100% nos dois contextos, esta diferença na

eficiência pode ser atribuída ao maior tempo que os usuários levaram para executá-la no contexto da câmera de documentos em relação ao contexto da câmera sem fio ($\alpha=5\%$, $F(3,32)=5,96$, $p=0,0061$).

Não houve diferença significativa entre a eficiência medida no contexto do emulador e no contexto da câmera sem fio em nenhuma das tarefas.

Estes resultados sugerem que o maior tempo para a execução das tarefas no contexto da câmera de documentos não pode ser atribuído ao uso do telefone celular, mas, provavelmente, foi ocasionado em função das condições de teste da câmera de documentos dentro do laboratório.

5.1.3 Eficiência relativa do usuário em relação ao usuário especialista

Para as tarefas 4 ($\alpha=5\%$, $F(3,32)=1,74$, $p=0,19$) e 7 ($\alpha=5\%$, $F(3,32)=0,57$, $p=0,56$) não houve diferença estatisticamente significativa entre as médias obtidas nos três contextos de avaliação (figura 51).

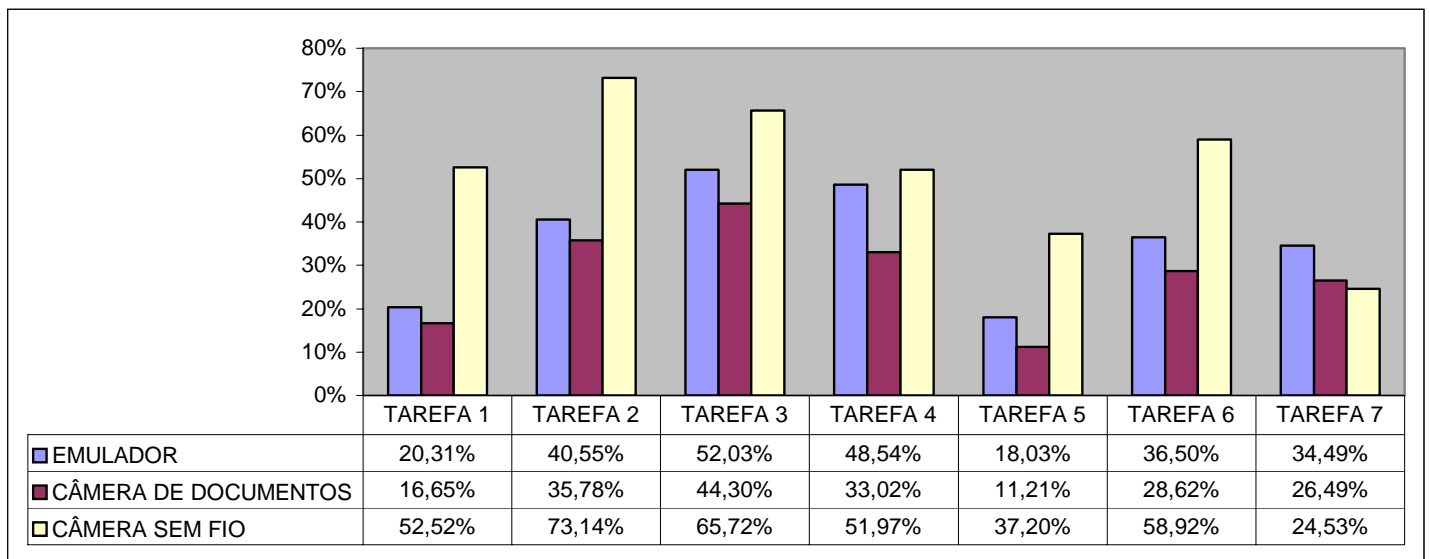


Figura 51: Resultados para a eficiência do usuário em relação ao usuário especialista

A eficiência relativa do usuário em relação ao usuário especialista para as tarefas 1, 2, 3, 5 e 6 foi maior no contexto da câmera sem fio em relação aos outros dois contextos de avaliação ($\alpha=5\%$, $F(3,32)=5,36$, $p=0,0096$; $F(3,32)=10,99$, $p=0,00021$; $F(3,32)=3,60$, $p=0,03$; $F(3,32)=5,89$, $p=0,0064$; $F(3,32)=7,62$, $p=0,0018$ respectivamente).

Entretanto nestas mesmas tarefas não houve diferença significativa para as médias nos contextos da câmera de documentos e do emulador.

Considerando que não foi constatada diferença significativa para a eficiência do usuário especialista nos três contextos de avaliação na realização das tarefas podemos concluir que, de maneira geral, os participantes tiveram um desempenho mais próximo do usuário especialista no contexto da câmera sem fio em relação aos outros dois contextos de avaliação.

Ou seja, a eficiência do usuário nos ensaios realizados dentro do laboratório ficou, de forma geral, bem abaixo da eficiência do especialista realizando as mesmas tarefas no mesmo local. Uma vez que não houve diferença entre as médias obtidas nos contextos do emulador e da câmera de documentos não é possível atribuir esse baixo desempenho ao telefone celular ou ao emulador. É possível concluir que o local de realização dos ensaios pode ter influenciado os participantes, pois no ambiente externo eles conseguiram um desempenho superior em relação ao laboratório. Isto indica que, possivelmente, a realização dos ensaios de interação fora do laboratório se mostrou como uma situação mais natural para todos.

5.1.4 Período produtivo

A análise das médias do período produtivo (figura 52) mostrou que houve diferenças estatisticamente significativas somente para a tarefa 1, onde o período produtivo foi menor no contexto do emulador em relação aos outros dois contextos de avaliação ($\alpha=5\%$, $F(3,32)=4,44$, $p=0,01$).

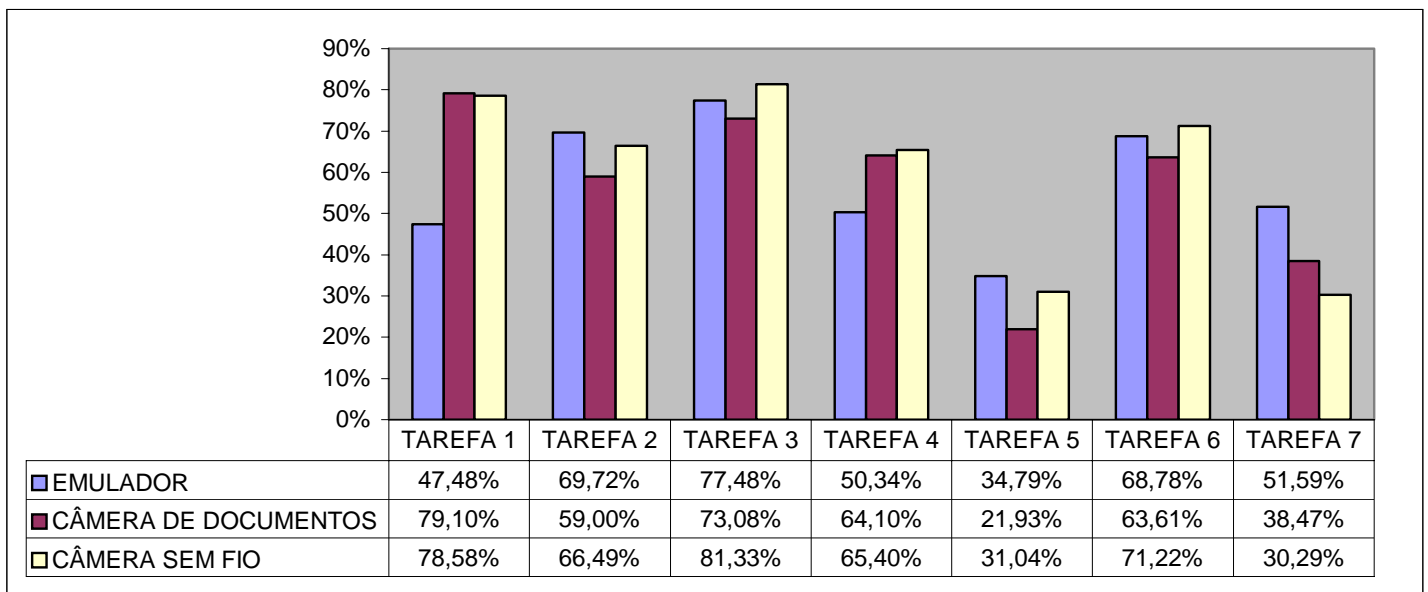


Figura 52: Resultados para o período produtivo

Para as demais tarefas não houve diferença estatisticamente significativa entre as médias dos tempos produtivos nos três contextos de avaliação ($\alpha=5\%$, $F(3,32)=1,08$, $p=0,35$; $F(3,32)=0,39$, $p=0,67$; $F(3,32)=1,79$, $p=0,18$; $F(3,32)=1,06$, $p=0,35$; $F(3,32)=0,40$, $p=0,67$; $F(3,32)=1,55$, $p=0,32$, para as tarefas 2 a 7 respectivamente).

O período produtivo foi menor na primeira tarefa para o contexto do emulador em relação aos outros dois contextos e não houve diferença significativa entre os contextos da câmera de documentos e da câmera sem fio para esta mesma tarefa.

Este resultado indica que na execução da primeira tarefa os participantes que utilizaram o emulador perderam mais tempo esbarrando em problemas em relação aos participantes que utilizaram o telefone celular nos outros dois contextos. Entretanto, essa diferença não ocorreu nas demais tarefas, ou seja, como a perda de tempo ocorreu somente para a primeira tarefa é possível concluir que possivelmente os participantes levaram um tempo um pouco maior para se acostumar com a interface do emulador. Todos os participantes eram usuários de computador e estavam habituados a navegar na Internet e, apesar de terem sido treinados no uso do emulador, nenhum deles havia utilizado uma interface similar anteriormente. A interface do telefone celular foi mais familiar aos participantes uma vez que todos já estavam habituados com o seu uso e mesmo havendo diferença entre os modelos, com o treinamento recebido eles foram capazes de mapear as teclas e funções que eles já estavam habituados a utilizar em seus próprios aparelhos.

5.1.5 Satisfação do usuário

O resultado da aplicação do questionário SUS (SUS, 2003) para medir a satisfação do usuário em relação à interação com o sistema nos três contextos de avaliação está ilustrado na figura 53.

Os resultados mostram que a satisfação do usuário em relação à interação com o sistema foi significativamente reduzida no contexto da câmera de documentos em relação aos outros dois contextos de avaliação (para $\alpha=5\%$, $F(3,32)=4,60$, $p=0,01$).

Estes resultados comprovam que as condições de realização dos ensaios de interação no contexto de avaliação da câmera de documentos impactaram negativamente sobre a usabilidade medida no que diz respeito à satisfação do usuário na interação com o sistema.

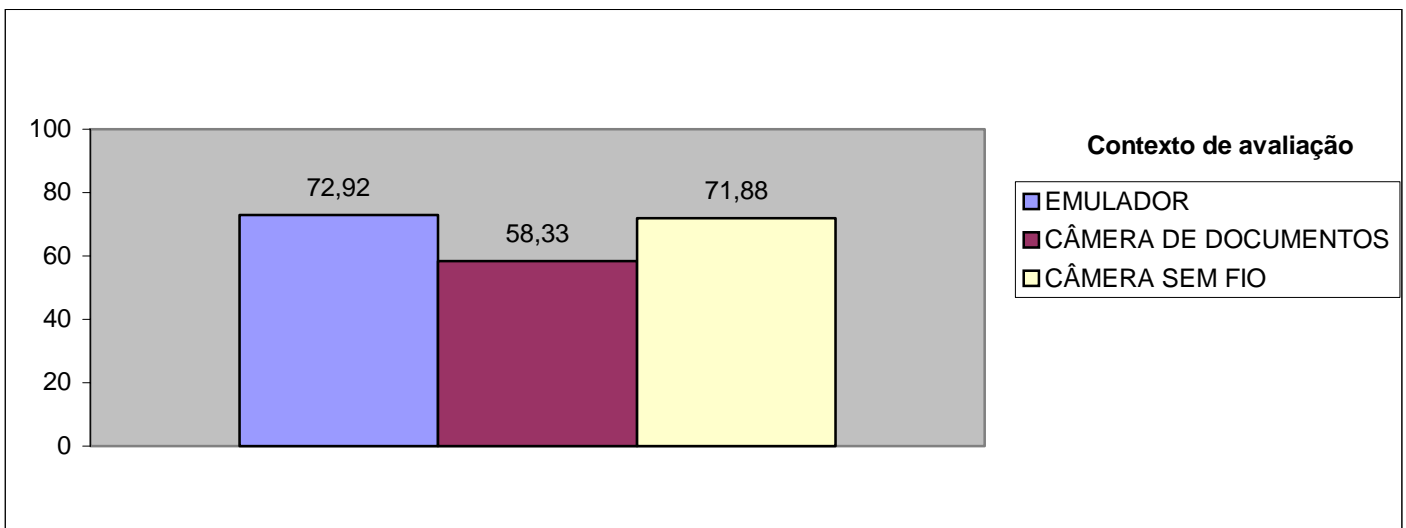


Figura 53: Resultados para o questionário SUS

5.1.6 Avaliação da carga de trabalho

A figura 54 mostra os resultados da aplicação do NASA TLX. Os resultados são apresentados para o valor médio geral da carga de trabalho em cada contexto de avaliação e para cada fator isoladamente, estes últimos obtidos a partir da média dos valores atribuídos pelos participantes a cada um dos seis fatores geradores de carga que foram analisados.

Quanto mais alto o valor apresentado por cada fator, maior é a sua contribuição na geração da carga de trabalho. É importante destacar que valores mais altos para a Performance indicam maiores graus de insatisfação do usuário em relação ao seu desempenho.

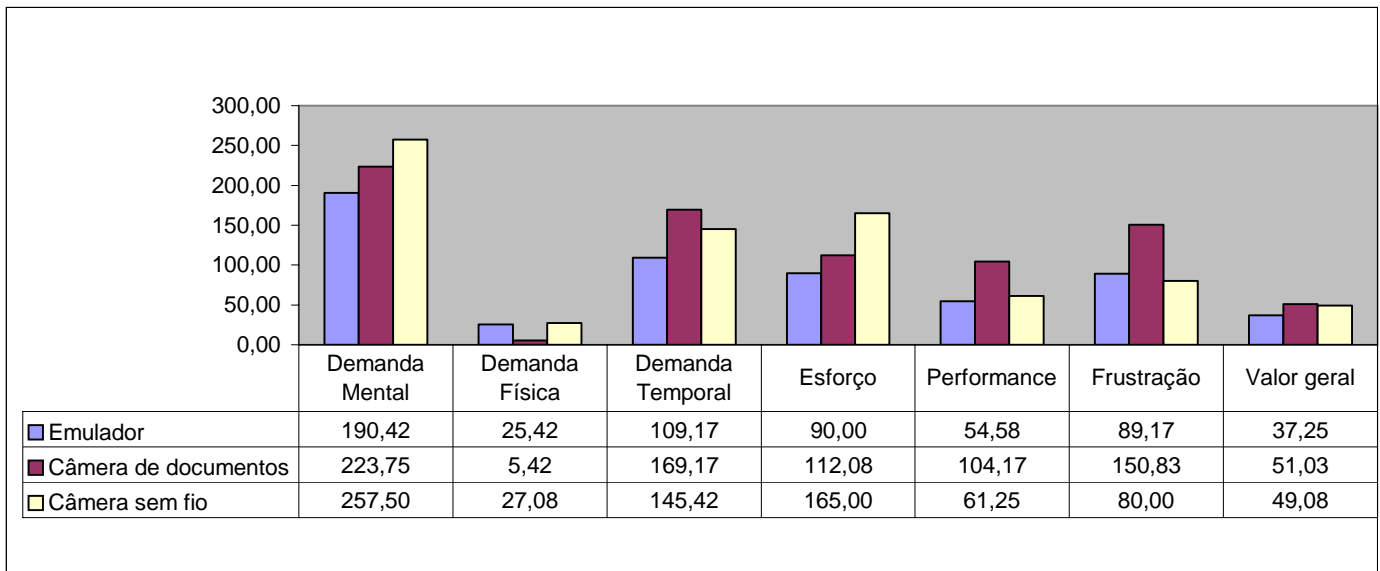


Figura 54: Resultados para o NASA TLX

Não houve diferenças estatísticas para os resultados obtidos para o valor geral do NASA TLX nos três contextos de avaliação ($\alpha=5\%$, $(F(3,32)=2,40, p=0,10)$).

Também não foi detectada diferença estatística entre os fatores isoladamente, a Demanda Mental ($\alpha=5\%$, $(F(3,32)=0,82, p=0,44)$), a Demanda Física ($\alpha=5\%$, $(F(3,32)=1,14, p=0,33)$), a Demanda Temporal ($\alpha=5\%$, $(F(3,32)=0,60, p=0,55)$), o Esforço ($\alpha=5\%$, $(F(3,32)=1,69, p=0,19)$), a Performance ($\alpha=5\%$, $(F(3,32)=2,31, p=0,12)$) e a Frustração ($\alpha=5\%$, $(F(3,32)=1,63, p=0,21)$).

Estes resultados indicam que sob o ponto de vista do usuário não houve diferença estatisticamente significativa entre os contextos de avaliação no que diz respeito à carga de trabalho.

5.2 Medidas e métricas de desempenho das avaliações

5.2.1 Quantidade de problemas de usabilidade

Considerando os três contextos de avaliação, foram reportados um total de 44 problemas de usabilidade na interação com o sistema. Dentre esses problemas um foi descartado por não corresponder a um problema real de usabilidade da interface do sistema avaliado pois ele foi ocasionado em função da interface do emulador. Este problema já havia sido mencionado em (Kaasinen, 2000) e é causado em função do emulador selecionado para as avaliações não apresentar uma indicação de que há mais texto além do que é visível na tela. Este problema ocorreu com dois participantes que, apesar de terem encontrado a tela correta, leram somente a parte do texto que estava visível na tela e não encontrando a informação procurada, voltaram à tela anterior. Somente após alguns segundos de navegação eles perceberam que era possível rolar a tela para baixo para encontrar o restante do texto e conseguiram completar a tarefa. Na tela do telefone celular há uma barra de rolagem no canto direito que indica quando há mais texto para cima ou para baixo. Todo problema de usabilidade deve estar relacionado a um ou mais aspectos da interface do sistema, que são considerados a causa dos problemas na interação. Este não é um problema que está relacionado a um aspecto da interface real do sistema uma vez que na tela do telefone celular esta indicação de continuidade do texto existe, e portanto não foi considerado como um problema de usabilidade real, mas sim provocado pela interface do emulador.

Após esta análise restaram 43 problemas de usabilidade reais, todos relacionados a aspectos da interface. Deste total, 35 problemas foram identificados

no contexto do emulador, 41 problemas foram identificados no da câmera de documentos e 38 problemas foram identificados no da câmera sem fio (figura 55).

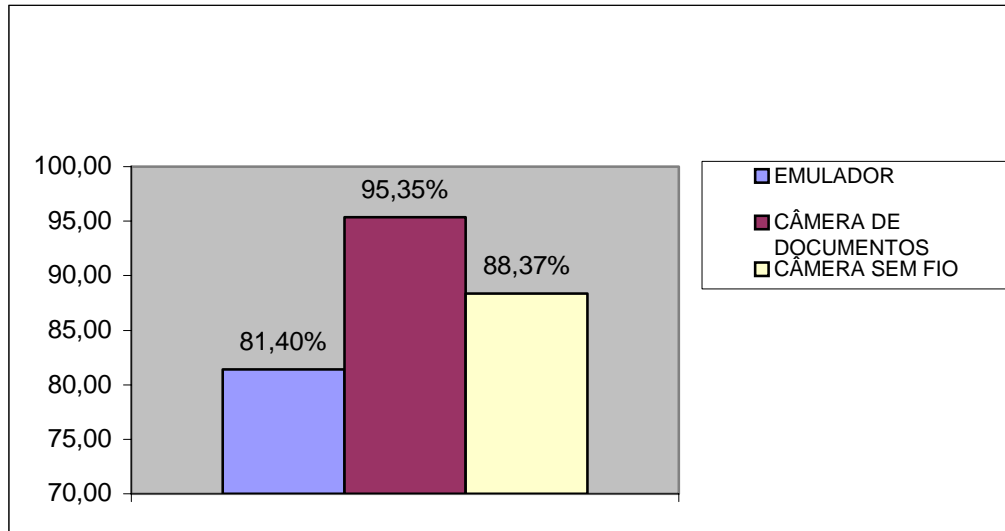


Figura 55: Percentual de problemas encontrados por contexto de avaliação

Estes resultados mostram que o contexto da câmera de documentos foi o contexto no qual foi possível identificar o maior número de problemas de usabilidade, ou seja, 95,35% de todos os problemas encontrados. Este valor corresponde a 8% a mais de problemas de usabilidade identificados em relação ao contexto da câmera sem fio e 17% a mais em relação ao contexto do emulador.

Considerando o número de problemas de usabilidade identificados, o contexto do emulador apresentou um resultado extremamente satisfatório uma vez que foi capaz de identificar mais de 80% do total de problemas.

Dentre os 43 problemas de usabilidade identificados, 32 foram comuns aos três contextos de avaliação. Os 11 restantes se distribuem entre os diferentes contextos de avaliação como ilustra a figura 56.

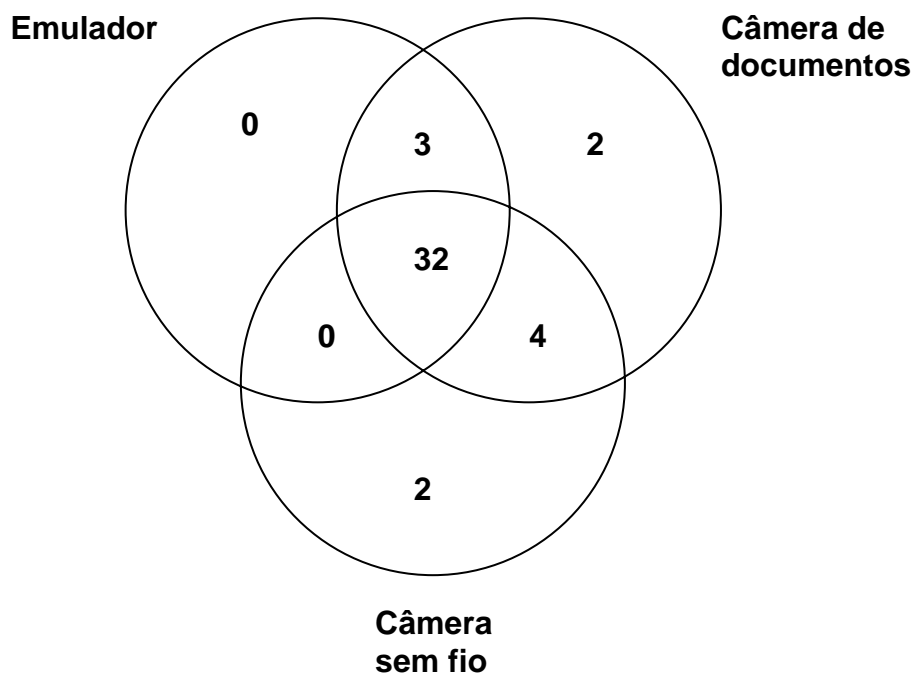


Figura 56: Número total de problemas de usabilidade distribuídos por contexto de avaliação

A partir da análise da distribuição dos problemas constata-se que não houve nenhum problema de usabilidade identificado exclusivamente no contexto do emulador, entretanto os contextos da câmera de documentos e da câmera sem fio detectaram 2 problemas exclusivos cada um.

Do total de 8 problemas que não foram identificados no contexto do emulador somente um deles era um problema relativo às características do aparelho de telefone celular que estava sendo utilizado e não poderia ter ocorrido no emulador. Este problema se refere à habilitação de um campo para entrada de texto. No emulador este campo já aparece habilitado enquanto que no telefone celular é necessário que o usuário pressione uma tecla antes de dar início à digitação. Como não há nenhuma indicação de que é necessário este passo a mais para habilitar o

campo, os participantes ficaram bastante confusos. A interface do emulador facilitou a usabilidade neste contexto uma vez que o campo já aparece habilitado.

5.2.2 Severidade dos problemas de usabilidade

A figura 57 mostra que o contexto da câmera de documentos foi o contexto que totalizou o maior grau de severidade de problemas tendo um resultado 23% superior em relação ao contexto da câmera sem fio e 59% superior em relação ao contexto do emulador. Entre os contextos da câmera sem fio do emulador o contexto da câmera sem fio apresentou um resultado 29% superior.

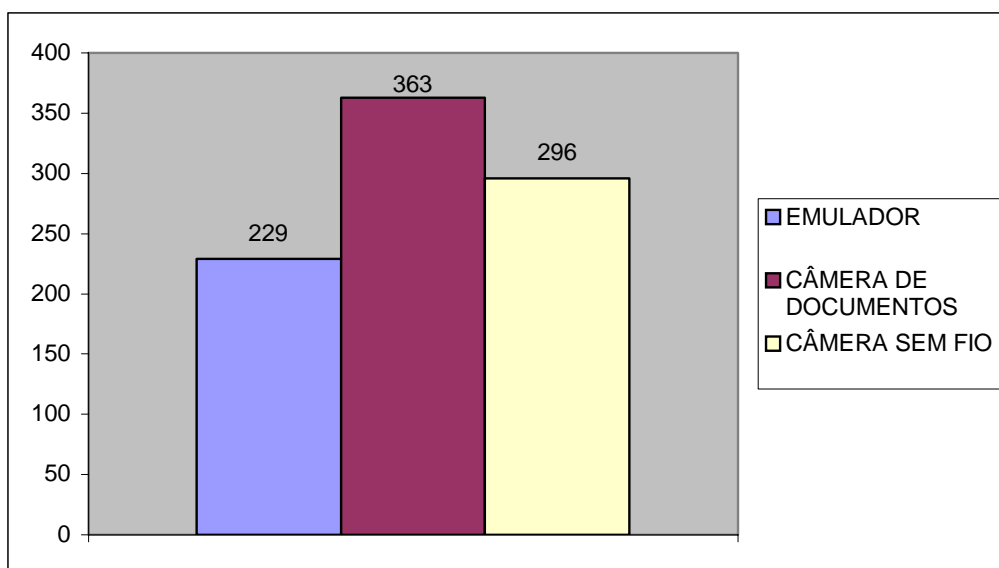


Figura 57: Total dos graus de severidade dos problemas de usabilidade encontrados por contexto de avaliação

A análise da figura 58 mostra que os problemas de usabilidade identificados ocorreram com uma frequência muito maior no contexto da câmera de documentos para todos os graus de impacto do problema.

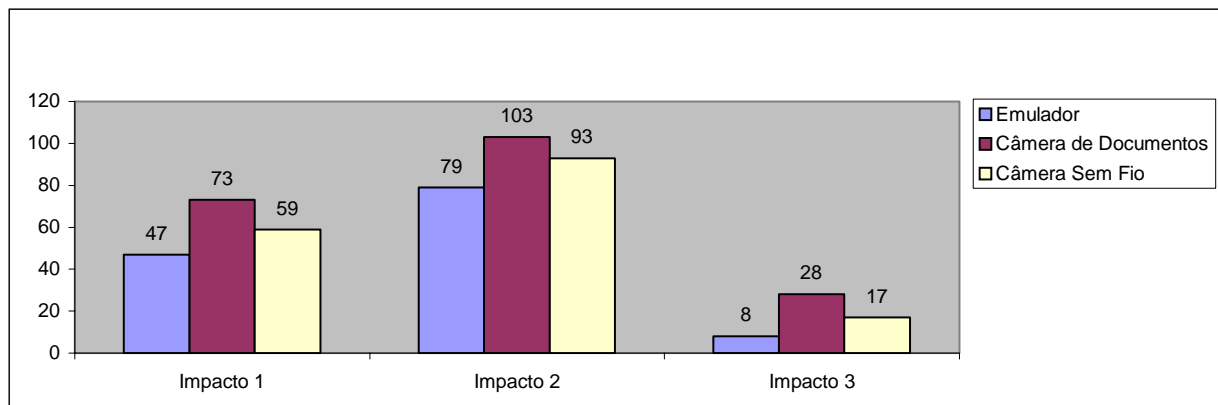


Figura 58: Frequência dos problemas de usabilidade distribuídos por impacto

Os problemas de impacto 1 ocorreram com uma frequência 55% maior no contexto da câmera de documentos em relação ao contexto do emulador e 24% maior em relação ao contexto da câmera sem fio. A frequência dos problemas de impacto 2 foi maior também no contexto da câmera de documentos sendo 24% superior em relação ao contexto do emulador e 11% superior em relação ao contexto da câmera sem fio. Para os problemas de impacto 3 esta diferença é altamente significativa, principalmente em relação ao contexto do emulador onde esta diferença foi de 250%, enquanto que para o contexto da câmera sem fio o valor foi de 65%.

A maior visibilidade dos problemas de impacto 3 no contexto da câmera de documentos é um aspecto importante a ser considerado uma vez que estes são considerados os problemas mais importantes a serem detectados por serem os mais graves, pois quando eles ocorrem o usuário é incapaz de atingir todos os seus objetivos na tarefa. Desta forma eles estão relacionados aos aspectos da interface que devem ser modificados prioritariamente, antes dos demais problemas.

A diferença entre a frequência de ocorrência dos problemas de impacto 3 no contexto do emulador e da câmera sem fio foi significativa, um valor de 113%,

indicando que os problemas mais graves aparecem com mais frequência no telefone celular. Para os problemas de impacto 1 e de impacto 2 esta diferença foi de 26% e 18% respectivamente.

Entretanto podemos considerar que apesar de aparecerem com menor frequência no contexto do emulador, este contexto foi capaz de identificar 75% dos problemas de maior impacto (impacto 2 e 3) enquanto que os contextos da câmera de documentos e da câmera sem fio identificaram 100% desses problemas.

5.2.3 Razão benefício/custo

A soma dos graus de severidade determina a eficácia de cada contexto de avaliação na detecção dos problemas de usabilidade. Como medida de custo foram consideradas as horas despendidas pelo avaliador na realização dos ensaios de interação e na análise das fitas de vídeo. Assim é possível determinar a eficiência em função de uma razão benefício/custo para cada um dos contextos que vai indicar a severidade dos problemas encontrados por hora de avaliação (figura 59).

A análise da razão benefício custo indica uma vantagem maior para o contexto da câmera de documentos, sendo 82% maior em relação ao contexto do emulador e 30% maior em relação ao contexto da câmera sem fio.

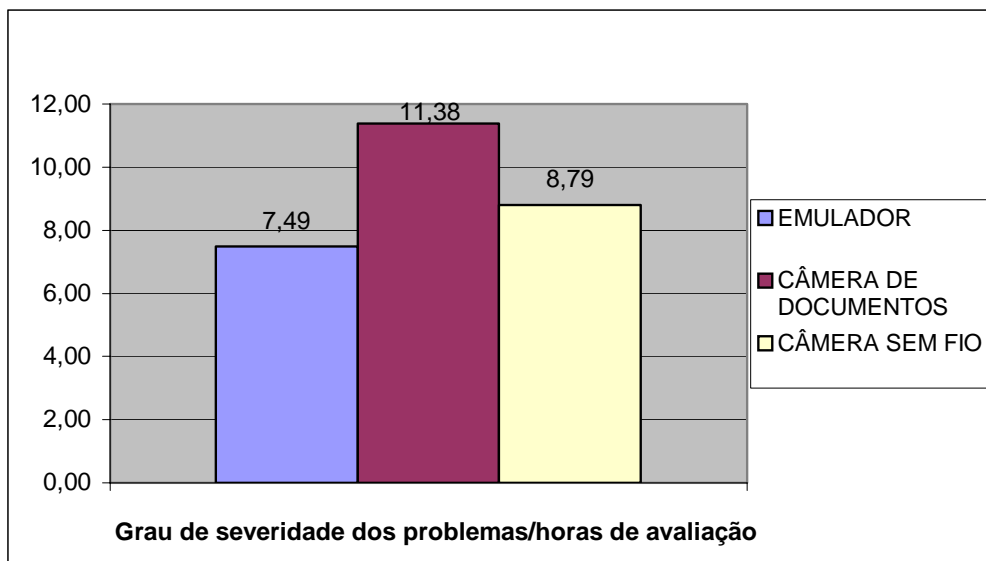


Figura 59: Razão benefício/custo por contexto de avaliação

5.3 Questionário 'Perfil do Usuário'

De acordo com os critérios estabelecidos para seleção dos usuários definido no item 4.3.5 do capítulo anterior, foram recrutados 36 participantes, 18 homens e 18 mulheres, distribuídos uniformemente por faixa etária entre 21 e 40 anos. Todos eram usuários de computador e telefone celular mas nunca haviam usado WAP anteriormente. Nenhum deles conhecia o telefone celular modelo LG BD-4000 (aparelho usado nos experimentos).

Além dessas características, a partir da análise das respostas ao questionário 'Perfil do Usuário' foi possível traçar um perfil mais detalhado sobre todos os participantes deste experimento.

Quanto ao grau de instrução 67% dos participantes possuíam nível superior completo ou mais, os restantes 33% possuíam nível superior incompleto (figura 60).

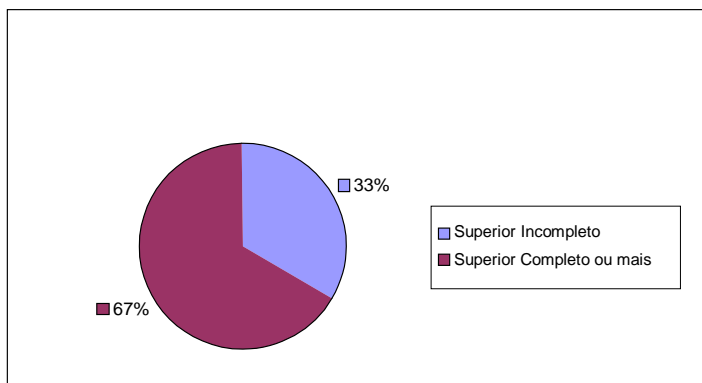


Figura 60: Grau de instrução dos participantes

Os 36 participantes eram estudantes e/ou profissionais de diversas áreas como: Direito, Administração, Marketing, Processamento de Dados, Fisioterapia, Jornalismo, Biologia, Recursos Humanos, Desenho Industrial, Educação Física, Relações Internacionais, Psicologia, Engenharia Civil, Engenharia Química, Engenharia Elétrica, Turismo, Comércio, Contabilidade, Pedagogia, Publicidade e Propaganda, Medicina, Engenharia Eletrônica, Assistência Social (figura 61).

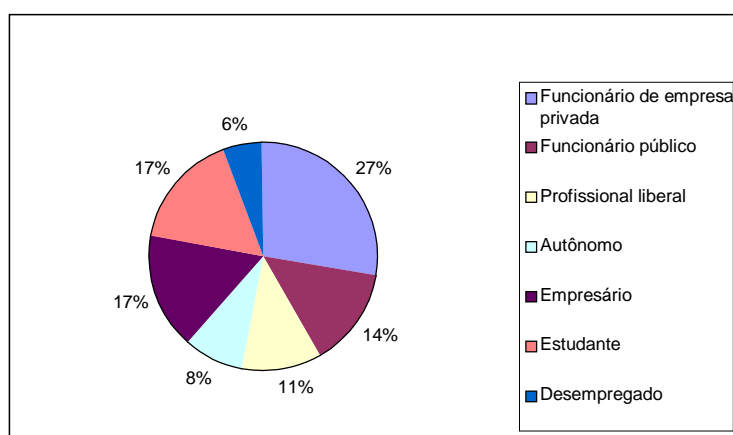


Figura 61: Ocupação dos participantes

Todos os participantes tinham no mínimo um ano de experiência com a Internet (figuras 62, 63, 64) e todos eram usuários de e-mail, sendo que a grande maioria, 94%, usava correio eletrônico há pelo menos um ano (figura 65).

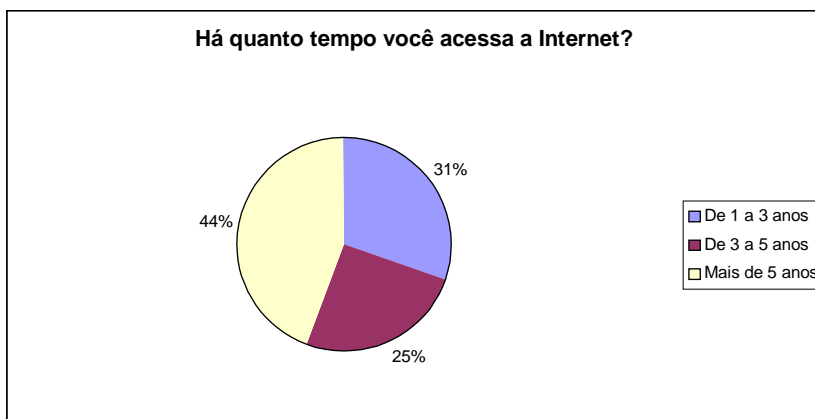


Figura 62: Tempo como usuário da Internet

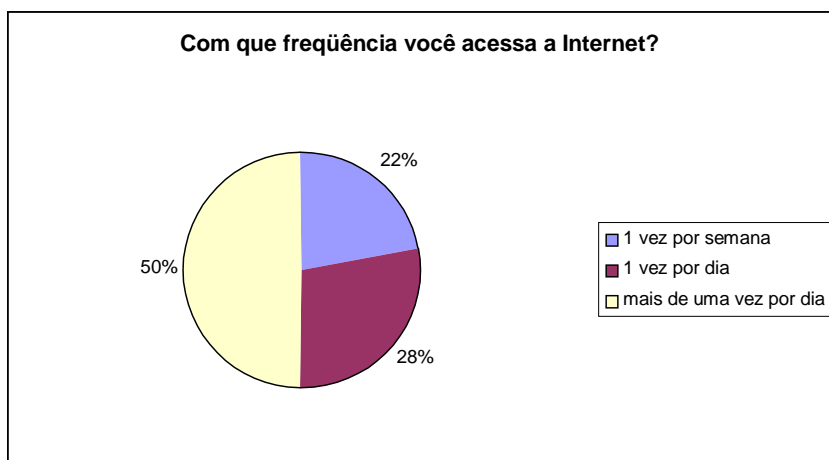


Figura 63: Frequência de acesso à Internet

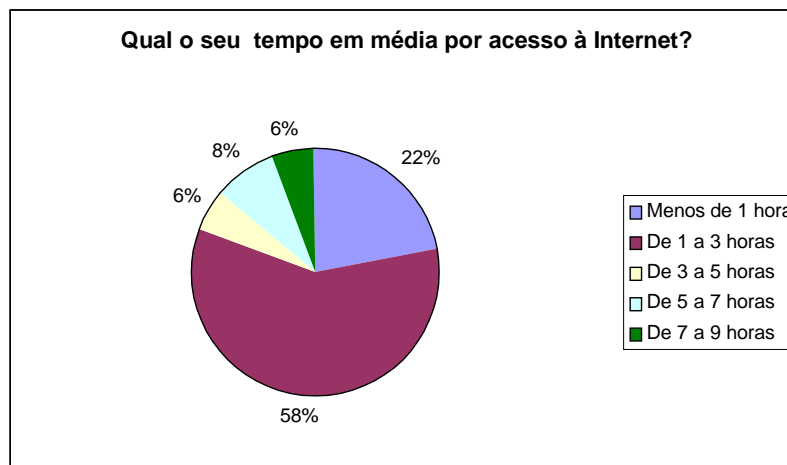


Figura 64: Média de tempo de acesso

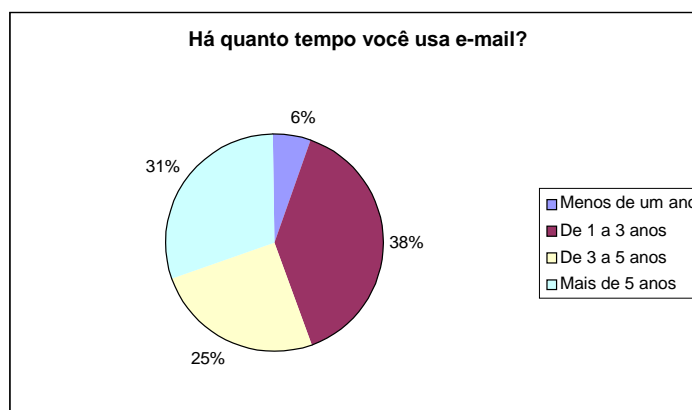


Figura 65: Tempo como usuário de e-mail

Todos eram usuários de telefone celular há pelo menos um ano (figura 66) e 64% disseram ter o costume de ler mensagens de texto no telefone celular no mínimo uma vez ao dia. Entretanto, o envio das mensagens é menos freqüente sendo que 30,56% disseram nunca ter enviado mensagens de texto pelo telefone celular (figura 67).

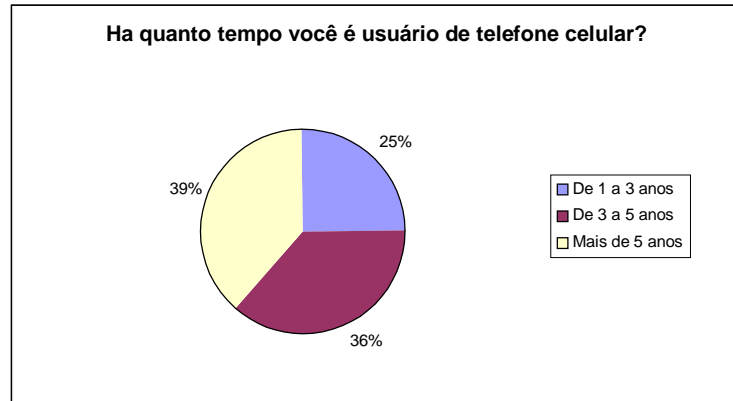


Figura 66: Tempo como usuário de telefone celular

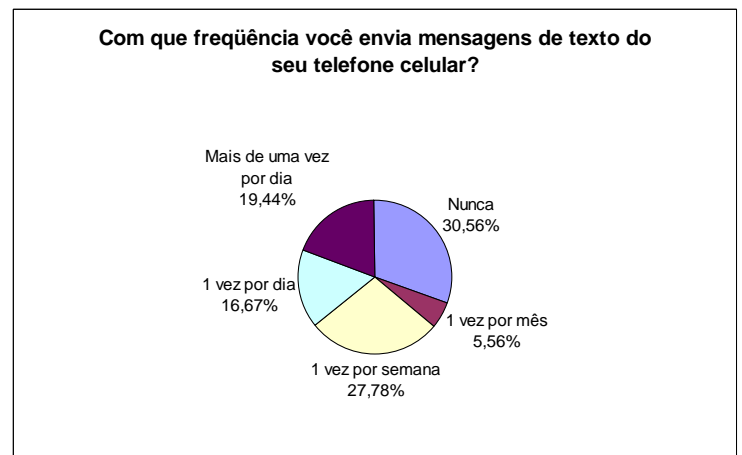
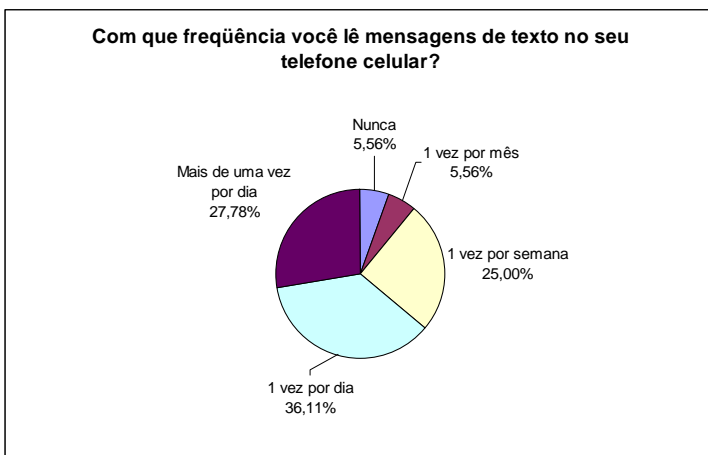


Figura 67: Frequência de leitura e envio de SMS

5.4 Questionário 'Avaliação geral sobre WAP'

Ao final dos ensaios de interação foi solicitado ao participante que respondesse a algumas perguntas para avaliar qual a sua percepção em relação à utilização dos serviços de Internet móvel. Os resultados deste questionário estão apresentados a seguir.

Entre os 36 participantes, 31% não sabiam dizer se o seu aparelho de telefone celular tinha acesso à Internet móvel e entre os 48% que sabiam que tinham telefones com este recurso nenhum havia utilizado esses serviços anteriormente. Estes dados provavelmente resultam da falta de informação sobre o que é este serviço, o que ele pode oferecer, quanto custa e como ele pode ser acessado. Um total de 87% dos participantes disseram não ter conhecimento sobre os diversos serviços oferecidos pela Internet móvel, mas de maneira geral todos os participantes ficaram surpresos positivamente em relação ao que estes serviços podem lhe oferecer. Os depoimentos de alguns dos participantes ao final do experimento em relação à sua experiência na navegação no portal WAP, ilustram essas constatações:

Participante 1: “Foi muito mais fácil do que eu imaginava, eu pensava que fosse muito mais complicado. Eu sabia que tinha esse recurso no meu telefone, mas eu não tinha a menor idéia de como funcionava”.

Participante 2: “Eu achei muito mais fácil e mais completo do que eu imaginava. Eu não sabia que era tão fácil, se não eu já teria utilizado antes”.

Participante 3: “Eu achei fácil e bem interessante. Você tem um acesso fácil, tá ali na tua mão, mas você acaba não usando por falta de informação. Eu nunca usei porque pensava que fosse caro e complicado”.

Participante 4: “Eu gostei, achei muito interessante mas eu nem sabia que essa tecnologia já estava disponível, e olha que eu me considero uma pessoa instruída e realmente não tinha conhecimento desse sistema”.

Apesar das limitações físicas dos telefones celulares em relação ao tamanho da tela e do teclado do aparelho, à primeira vista parece que os participantes não vêem o tamanho reduzido das interfaces físicas como um empecilho a utilizar um telefone celular para navegar na Internet móvel (figuras 68 e 69).

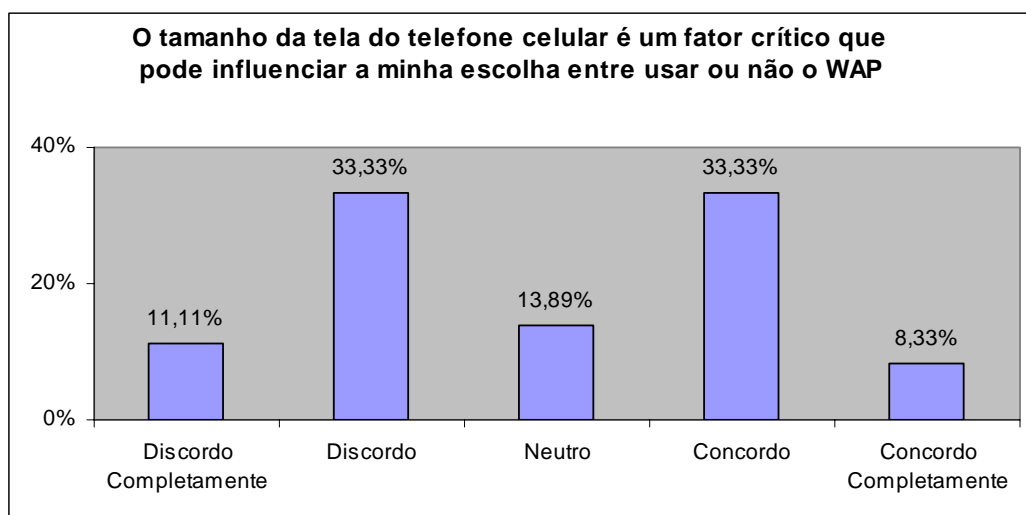


Figura 68: Opinião sobre o tamanho da tela do telefone celular

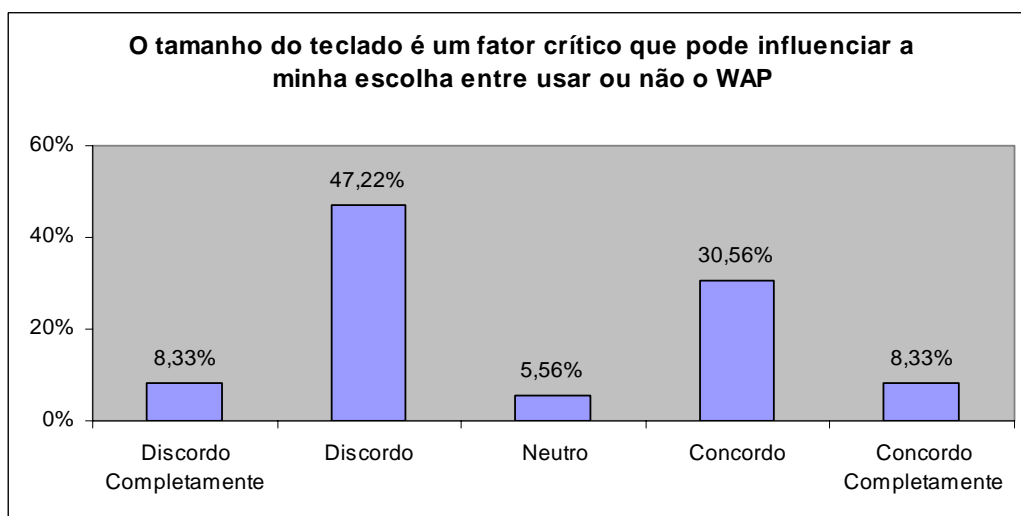


Figura 69: Opinião sobre o tamanho do teclado do telefone celular

Na realidade 89% dos participantes acreditam que se eles utilizarem o seu próprio aparelho de telefone celular, com o qual eles já estão habituados ao tamanho da tela e às funcionalidades do teclado, a navegação não será um problema (figura 70).

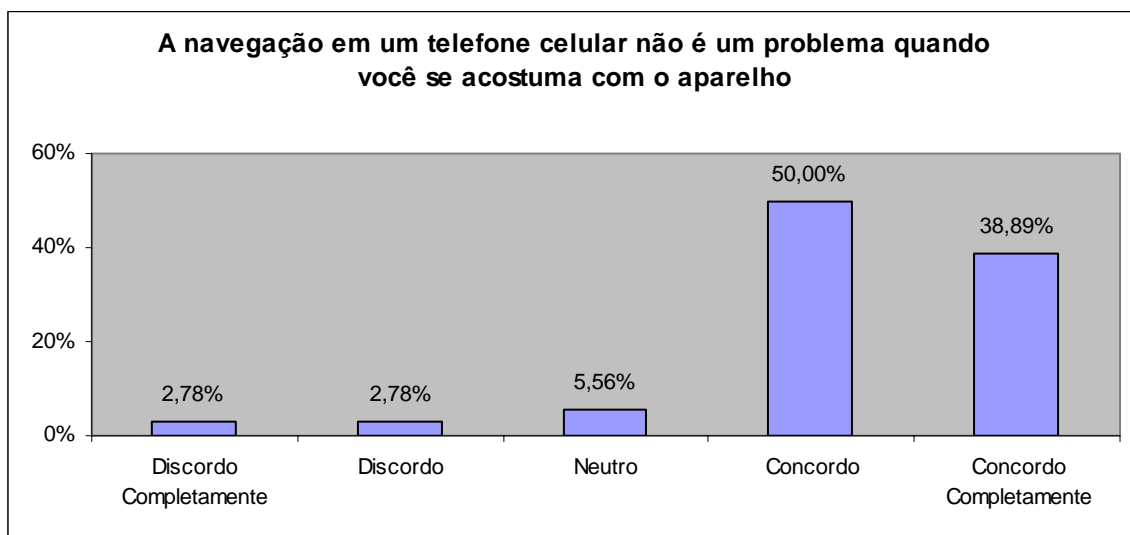


Figura 70: Opinião sobre a navegação em um telefone celular

Entretanto todos são unânimes em afirmar que uma conexão estável e rápida é fator determinante para que eles possam utilizar esses serviços como mostra a figura 71.

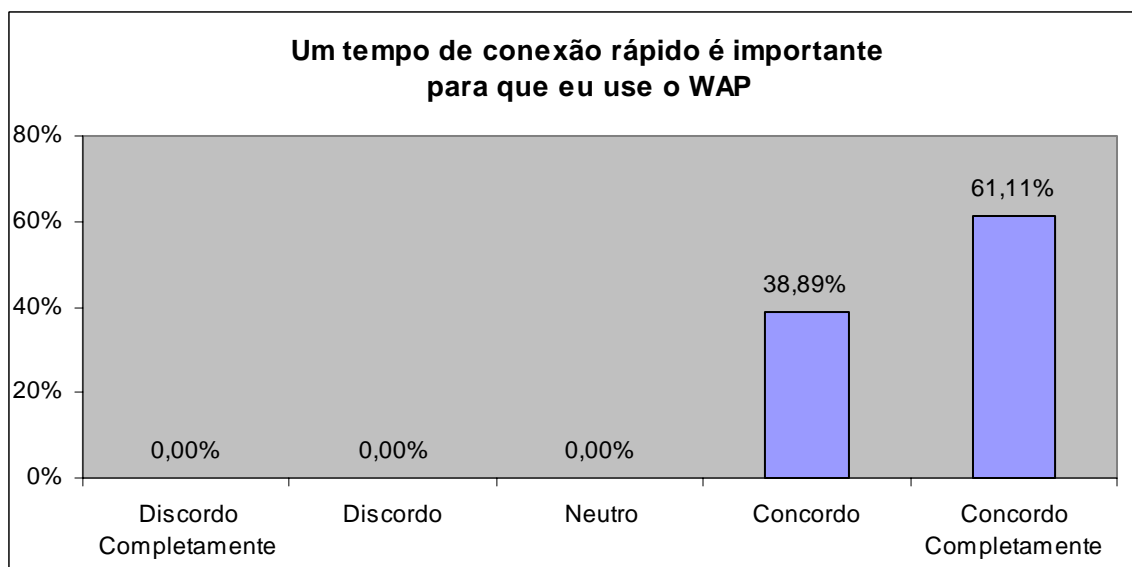


Figura 71: Opinião sobre o tempo de conexão

A qualidade do conteúdo oferecido também foi considerado um fator importante para a maioria dos participantes como mostra a figura 72.

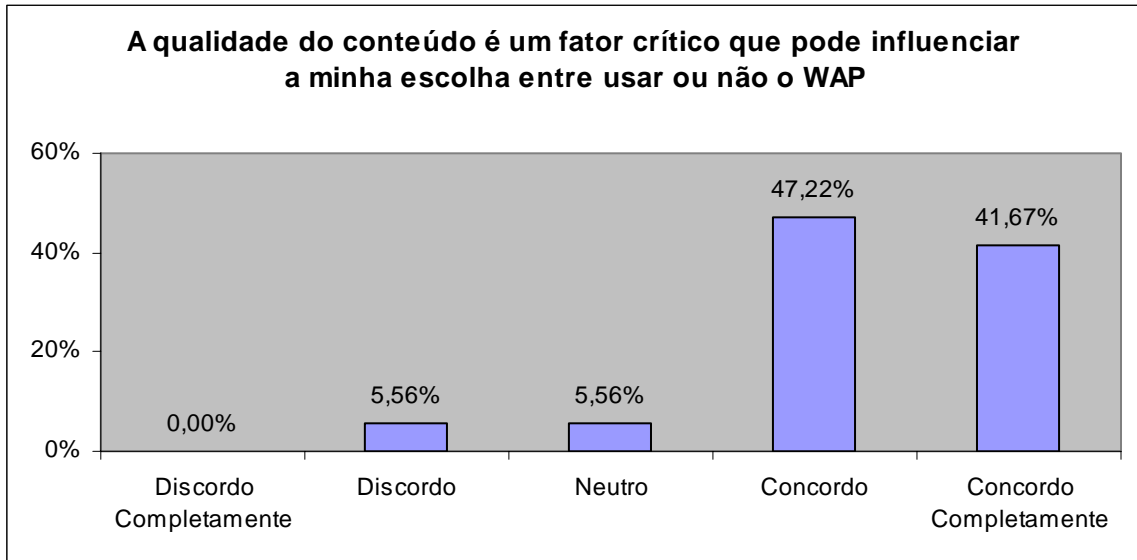


Figura 72: Opinião sobre a qualidade do conteúdo

Embora o custo de acesso desses serviços seja um fator preocupante para a maioria dos participantes como mostra a figura 73, muitos concordam que esse custo pode ser compensado em função da utilidade e da praticidade que o serviço pode oferecer ao usuário móvel. O depoimento de um participante ilustra esta perspectiva: “Eu coloco o custo em um contexto de negócio: se eu estiver no aeroporto, pronto para pegar o avião e precisar enviar um e-mail importante, eu vou usar o serviço, independente de quanto ele custar”. Este mesmo participante completa o seu raciocínio comprovando o que já foi constatado por (Nielsen, 2000) e (Helyar, 2002) em relação aos serviços de Internet móvel concorrerem com outras mídias como jornais, revistas e a própria Internet: “Eu pagaria para mandar esse e-mail, mas eu não pagaria para ver a programação de teatro no Rio de Janeiro. Para isso eu compraria um jornal”.

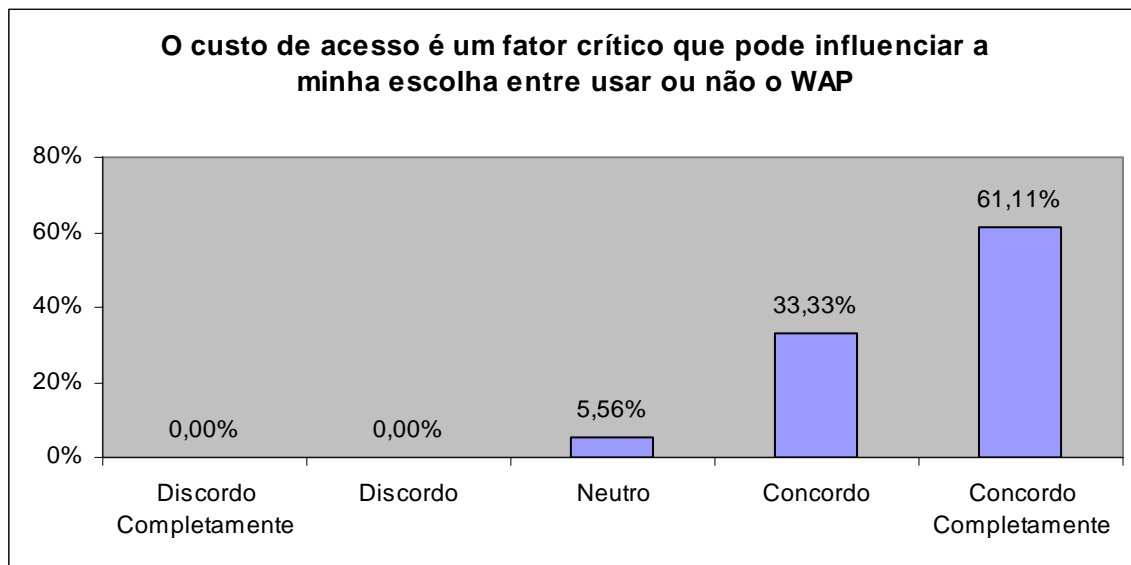


Figura 73: Opinião sobre o custo de acesso

De uma forma geral a reação dos participantes à idéia de utilizar um telefone celular para navegar na Internet móvel foi bastante positiva, como declarou um dos participantes: “Eu vou sair daqui curioso para saber dos outros serviços, o que eles oferecem e como eles funcionam”.

6 CONCLUSÃO

No experimento realizado os ensaios de interação foram conduzidos segundo três diferentes abordagens caracterizadas pela definição dos seus contextos de avaliação.

A seguir serão apresentadas a análise e a discussão dos resultados obtidos neste experimento com o objetivo de esclarecer qual foi a influência exercida pelos diferentes contextos de avaliação nos resultados dos ensaios de interação.

No final serão apresentadas as principais conclusões deste trabalho, bem como sugestões para trabalhos futuros.

6.1 Análise dos resultados

Os resultados obtidos podem ser analisados em dois grupos:

- Os resultados da usabilidade medida em termos de eficácia, eficiência e satisfação do usuário incluindo a avaliação da carga de trabalho.
- Os resultados das medidas e métricas de desempenho das avaliações em termos de número de problemas de usabilidade identificados, severidade desses problemas e número de horas gastas na avaliação.

6.1.1 Usabilidade medida nos três contextos de avaliação

O quadro 5 apresenta um resumo da comparação dos resultados da usabilidade medida nos três contextos de avaliação, indicando onde ocorreu uma diferença estatisticamente significativa para esses resultados.

Quadro 5: Resumo das diferenças significativas na usabilidade medida nos três contextos de avaliação

	Câmera de documentos X Emulador							Câmera de documentos X Câmera sem fio							Câmera sem fio X Emulador							
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7	
Satisfação	≠ ¹							≠ ¹							=							
Carga de Trabalho	=							=							=							
	Tarefa							Tarefa							Tarefa							
Eficácia	≠ ¹	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	=	
Eficiência	=	≠ ¹	≠ ¹	=	≠ ¹	≠ ¹	=	=	≠ ¹	=	=	=	=	=	=	=	=	=	=	=	=	
Período Produtivo	≠ ²	=	=	=	=	=	=	=	=	=	=	=	=	=	≠ ²	=	=	=	=	=	=	
Eficiência Relativa	=	=	=	=	=	=	=	≠ ³	≠ ³	≠ ³	=	≠ ³	≠ ³	=	≠ ³	≠ ³	=	=	=	=	≠ ³	=

¹ menor no contexto da CÂMERA DE DOCUMENTOS

² menor no contexto do EMULADOR

³ maior no contexto da CÂMERA SEM FIO

6.1.1.1 Eficácia, eficiência e satisfação

Verifica-se que as medidas de eficácia e eficiência, para algumas tarefas, foram menores no contexto da câmera de documentos em relação aos outros dois contextos de avaliação. Os resultados da avaliação de satisfação do usuário em relação à interação com o sistema também foram significativamente

menores no contexto da câmera de documentos. Ou seja, o contexto da câmera de documentos impactou negativamente na usabilidade medida em relação aos outros dois contextos de avaliação.

Por outro lado, de uma maneira geral, não houve diferença estatisticamente significativa na usabilidade medida quando foram comparados os resultados do contexto do emulador com o contexto da câmera sem fio. Ou seja, os fatos de utilizar interfaces diferentes como o emulador e o telefone celular, e de realizar os testes dentro do laboratório e em campo, não alteraram os resultados da usabilidade medida nestes dois contextos.

Desta forma, não se pode atribuir os menores resultados para a usabilidade medida no contexto da câmera de documentos às diferenças existentes entre as interfaces do emulador e do telefone celular ou ao local de realização dos ensaios de interação, mas sim às condições de realização do teste com a câmera de documentos dentro do laboratório. Essas condições se referem à posição do telefone celular fixado sobre o tripé e posicionado sob a câmera de documentos.

Um desconforto foi causado principalmente pelo fato do telefone estar fixado em um tripé móvel, mas que não poderia ser movimentado pelo usuário, sob pena de sair da área de visualização da câmera de documentos. Para o participante esta não foi uma situação normal de manipulação do aparelho. O fato de precisar operar o telefone que estava posicionado sobre o tripé aliado à preocupação em manter a tela do telefone em uma posição favorável para a filmagem não pareceu ser uma condição muito natural aos participantes e

provocou um desconforto no usuário que acabou por influenciar o resultado dos testes.

6.1.1.2 Período produtivo e eficiência relativa do usuário

Ao analisar o período produtivo verifica-se que não houve diferença significativa alguma entre os três contextos de avaliação, exceto para a primeira tarefa que teve um período produtivo menor no contexto do emulador em relação aos outros dois contextos. Esta queda no período produtivo para a primeira tarefa no emulador revelou que a interface do telefone na tela do computador, apesar de parecer intuitiva, é uma interface virtual e que o usuário precisa de um pouco mais de tempo para se acostumar com a mesma. Esta diferença não ocorreu com o telefone celular indicando que, mesmo não conhecendo o modelo de telefone usado nos testes, o usuário acostumado ao uso de telefones celulares se adaptou rapidamente ao aparelho, não apresentando dificuldades no manuseio das teclas.

Os melhores resultados obtidos para a eficiência relativa do usuário no contexto da câmera sem fio em relação aos outros dois contextos indicam que o ambiente externo se mostrou mais favorável ao participante uma vez que neste contexto a sua eficiência ficou mais próxima à eficiência do usuário especialista.

6.1.1.3 Carga de trabalho

Antes de iniciar o experimento supunha-se que o contexto da câmera sem fio apresentaria os maiores índices de carga de trabalho, uma vez que o usuário estava exposto a um ambiente externo, que exigiria maior concentração nas tarefas em relação ao ambiente do laboratório. Entretanto, durante a realização dos ensaios, o avaliador percebeu que os usuários que participaram dos ensaios no contexto da câmera sem fio estavam completamente à vontade no ambiente externo, concentrados nas tarefas que estavam executando, sem enfrentar maiores dificuldades na interação e apresentando índices de usabilidade medida, em muitas oportunidades, superiores ao obtidos nos ensaios de interação realizados dentro do laboratório. Esta constatação é confirmada pelos resultados da avaliação da carga de trabalho (NASA TLX), que não mostraram diferenças significativas nos três contextos de avaliação.

Assim, o ambiente externo não foi responsável por um maior índice de carga de trabalho, ao mesmo tempo em que se pôde constatar que o ambiente da câmera de documentos, mesmo com índices menores para a usabilidade medida, não foi visto pelo participante como maior gerador de carga de trabalho. Ou seja, o desconforto causado pelo tripé e a câmera de documentos, foi uma tarefa cognitiva concorrente que, apesar de ter contribuído para diminuir os índices da usabilidade medida para os usuários no contexto da câmera de documentos, não foi percebida pelo usuário como uma maior fonte de carga de trabalho.

Estes resultados indicam que houve um equilíbrio entre os contextos de avaliação no sentido de que todos foram vistos pelos participantes da mesma maneira em termos da influência dos diversos fatores na geração da carga de trabalho. Assim, sob a perspectiva dos participantes, a experiência subjetiva da carga de trabalho foi a mesma nos três contextos de avaliação

6.1.2 Desempenho das avaliações nos três contextos de avaliação

Em relação às medidas e métricas de desempenho obtidas pelas avaliações, o contexto da câmera de documentos foi o que produziu os melhores resultados, como ilustra a tabela 2. Este contexto proporcionou a identificação do maior número de problemas de usabilidade, sendo que nele todos os problemas, principalmente os de maior impacto, apareceram com uma frequência maior em relação aos outros dois contextos de avaliação.

Tabela 2: Resumo das medidas e métricas de desempenho das avaliações

	Emulador 3º	Câmera de documentos 1º	Câmera sem fio 2º
Percentual de problemas de usabilidade identificados	81,4%	95,35%	88,37%
Severidade dos problemas	229	363	296
Benefício/custo	7,49	11,38	8,79

Considerando as horas gastas na avaliação, a câmera de documentos foi o contexto que apresentou a melhor relação benefício/custo uma vez que foi capaz de identificar o maior número de problemas, incluindo os mais graves, por hora de avaliação.

O contexto do emulador obteve ótimos índices de desempenho na identificação de problemas de usabilidade, considerando que neste contexto foi possível identificar 81,4% dos problemas. Este contexto também foi capaz de identificar 75% dos problemas de impacto 3, considerados os problemas de maior gravidade, embora estes problemas tenham aparecido com uma frequência muito menor em relação aos outros dois contextos de avaliação.

6.2 Discussão

O contexto do emulador é a situação de avaliação mais simples de ser executada uma vez que não exige equipamentos especiais para registrar a interação e não há a necessidade de utilizar o equipamento real no qual a aplicação será utilizada. A pertinência dos problemas de usabilidade identificados no contexto do emulador depende da similaridade entre a interface do emulador e a interface do telefone real utilizado. As pequenas diferenças existentes nas interfaces escolhidas neste experimento mostraram que uma leve mudança na interface do emulador, como um maior número de linhas por exemplo, pode causar uma repercussão muito grande na usabilidade verificada, para melhor ou para pior.

Ao mesmo tempo em que foi a situação que apresentou os melhores resultados para a avaliação, o contexto da câmera de documentos foi para o participante a situação mais desfavorável. Isto se deve, não só à questão do local do laboratório, mas principalmente aos equipamentos utilizados para registrar a interação. Apesar de todos os participantes terem sido unânimes em

afirmar que o tripé e a câmera de documentos não atrapalharam a realização das tarefas, os resultados indicam que o simples fato de ter um tripé não solidário à câmera de documentos foi uma fonte de estresse que teve influência na usabilidade medida.

A solução da minicâmera sem fio para registrar a avaliação no ambiente externo foi extremamente satisfatória uma vez que foi capaz de gerar imagens de alta qualidade sem prejudicar o desempenho do usuário. De uma maneira geral, os resultados da usabilidade medida no contexto da câmera sem fio estiveram próximos aos resultados obtidos nos outros dois contextos de avaliação, pois não houve diferença significativa na comparação desses resultados. Exceto para a eficiência relativa, cujos resultados mostraram que a eficiência dos usuários no contexto da câmera sem fio esteve mais próxima à eficiência dos usuários especialistas. Estas constatações podem ser atribuídas a dois fatores principais.

O primeiro se refere à posição dos participantes, que realizaram os ensaios de interação sentados (e não em pé, como se poderia configurar um possível contexto real), tanto nos ensaios no laboratório quanto nos ensaios realizados no ambiente externo. O outro fator decorre do fato de que as pessoas que possuem telefone celular não se sentem inibidas em utilizar o telefone em público. Elas já estão acostumadas a ignorar o ambiente e se concentrar na interação. O fato de que elas não estavam em uma situação totalmente natural, pois sabiam que estavam participando de uma avaliação e que o avaliador estava próximo, também pode ter ajudado a aumentar a concentração nas tarefas que estavam sendo executadas.

De maneira geral, a minicâmera acoplada ao telefone não prejudicou a manipulação do aparelho. Os participantes se sentiram à vontade no contexto da câmera sem fio, que proporcionou uma situação mais próxima da realidade de uso deste tipo de equipamento.

Analisando os contextos de avaliação sob a perspectiva do avaliador, o contexto da câmera sem fio se mostrou como o ambiente mais desconfortável para a realização dos ensaios de interação. O laboratório de testes é um ambiente mais confortável para o avaliador, é um local seguro, sob controle e com o qual o avaliador já está acostumado. É um local silencioso, onde todo o ambiente já está preparado para as avaliações, com todos os equipamentos prontos e conectados. O ambiente externo é um local novo e totalmente imprevisível. Neste ambiente o avaliador estava mais exposto a interferências externas que praticamente não existem dentro do laboratório. Como o ambiente era público e bastante movimentado, a qualquer momento uma pessoa poderia se aproximar para saber o que estava acontecendo, interrompendo a avaliação. É muito grande também a preocupação do avaliador não só em relação à segurança e ao bem estar do participante, mas também em relação aos equipamentos. Além da necessidade de instalar todos os equipamentos no momento do teste, a minicâmera, o receptor e a filmadora funcionam a bateria, o que implica na necessidade de levar unidades sobressalentes caso alguma bateria apresente algum problema durante o teste. Apesar das baterias utilizadas terem autonomia suficiente para registrar a interação durante uma hora de teste, esta foi uma fonte de preocupação

constante para o avaliador. Todos esses fatores podem interferir no seu nível de atenção e concentração.

O quadro 6 apresenta um resumo da análise da comparação dos resultados obtidos nos três contextos de avaliação.

Quadro 6: Resumo da comparação dos resultados nos três contextos de avaliação

	Contexto do Emulador	Contexto da Câmera de Documentos	Contexto da Câmera Sem Fio
Pontos positivos	<ul style="list-style-type: none"> - Ambiente controlado. - É o contexto de avaliação mais simples, pois não exige equipamentos especiais para registrar a interação e não é necessário possuir o aparelho real. 	<ul style="list-style-type: none"> - Ambiente controlado. - A situação de teste favoreceu o aparecimento dos problemas de usabilidade. - Melhor razão benefício/custo considerando custo como horas de avaliação. 	<ul style="list-style-type: none"> - É o ambiente mais natural para o usuário. - Maior flexibilidade para manusear o telefone celular.
Pontos negativos	<ul style="list-style-type: none"> - As pequenas diferenças existentes entre as interfaces do emulador e do equipamento real podem Ter grande impacto na usabilidade tanto para melhor quanto para pior. - É necessário que o participante seja usuário de computador. 	<ul style="list-style-type: none"> - É o ambiente mais desfavorável para o usuário. 	<ul style="list-style-type: none"> - Ambiente imprevisível. - É o ambiente mais desfavorável para o avaliador.

Este estudo mostrou que, nas condições em que os ensaios de interação foram realizados neste experimento, foi possível identificar o maior número de problemas de usabilidade dentro do laboratório, no contexto da câmera de documentos.

Mesmo para o contexto do emulador os resultados obtidos foram satisfatórios. É importante salientar que a interface escolhida para o emulador foi a mais semelhante possível à interface do equipamento real. Garantida esta condição, este contexto pode fornecer bons resultados a um custo baixo em termos de horas e de equipamentos necessários para registrar a avaliação.

O contexto da câmera sem fio, tal como preparado neste trabalho, mostrou-se particularmente indicado para auxiliar a reproduzir possíveis situações reais de uso deste tipo de equipamento. O sistema envolvendo a minicâmera sem fio e o suporte, desenvolvido especialmente para esta aplicação, mostrou ser uma excelente solução para auxiliar na realização de ensaios de interação sempre que for necessário registrar as imagens e o áudio da interação do usuário com um computador de mão. A mobilidade do usuário, aliada à liberdade de manipulação do telefone celular, torna possível a utilização deste sistema em qualquer ambiente, inclusive dentro do laboratório.

6.3 Conclusão

A pesquisa bibliográfica realizada indicou que as avaliações de usabilidade para os computadores de mão não devem ser conduzidas da maneira tradicional dentro do laboratório, uma vez que neste ambiente não seria possível reproduzir as condições de uso deste tipo de equipamento (Johnson, 1998; Petrie, 1998; Waterson, 2002; Holmquist, 2002; Brewster, 2002). Alguns autores, como Chittaro (2001), Pearrow (2002) e Vyas (2002), atestam que o

uso de emuladores nas avaliações de usabilidade dos computadores de mão não podem fornecer indicações confiáveis para a usabilidade medida.

A partir destas constatações foi formulada a hipótese básica deste trabalho: “Na avaliação de usabilidade dos computadores de mão os ensaios de interação realizados dentro do laboratório não fornecem resultados tão satisfatórios quanto os resultados obtidos nos ensaios de interação realizados fora do laboratório, com o usuário em uma situação mais próxima ao contexto de uso deste tipo de equipamento”.

Os resultados deste trabalho mostraram que, nos contextos específicos em que o experimento aqui relatado foi conduzido, a hipótese formulada pode ser rejeitada. Ao comparar os resultados das avaliações conduzidas dentro do laboratório, nos contextos do emulador e da câmera de documentos, em relação às avaliações conduzidas no ambiente externo, no contexto da câmera sem fio, não foi identificada uma diferença significativa em termos da usabilidade medida e dos índices de desempenho das avaliações. Esta constatação pode levar a duas conclusões.

A primeira conclusão indica que é possível obter bons resultados na avaliação de usabilidade de uma interface de um computador de mão usando um emulador no computador, desde que o contexto de avaliação seja definido da maneira mais próxima possível ao contexto de uso. Além de selecionar usuários representativos do público alvo e tarefas representativas das tarefas utilizadas normalmente por esses usuários, a escolha do emulador tem um papel fundamental. Foi verificado que pequenas diferenças na interface do emulador podem afetar a usabilidade. Desta forma, este experimento constatou

que bons resultados com o uso de um emulador podem ser obtidos desde que fique assegurada a similaridade entre as interfaces do emulador e do telefone celular, tanto na aparência quanto no comportamento.

Entretanto o uso do emulador coloca uma restrição na escolha da amostra de participantes, uma vez que é preciso que o participante já seja usuário de computador. O treino no uso do mouse e de outros dispositivos para a interação é importante, caso contrário as dificuldades com o uso do computador serão as principais enfrentadas pelo usuário, impossibilitando a avaliação do diálogo com a interface (Nielsen, 1993a). Neste experimento todos os participantes eram usuários com pelo menos um ano de experiência no uso da Internet. Mesmo assim, a medida do período produtivo mostrou que eles apresentaram dificuldades iniciais com o uso da interface do emulador, apesar do treinamento recebido no início da sessão. Por outro lado, o usuário de um computador de mão não precisa ser necessariamente um usuário de computador. Ou seja, a realização de outro experimento verificando o comportamento do usuário novato no uso do computador forneceria novos resultados, não só para determinar o seu comportamento face à interface do emulador mas também em relação à navegação na aplicação de Internet móvel, que segue um modelo hierárquico de menus já bem conhecido do usuário de Internet.

A segunda conclusão importante se refere à realização dos ensaios no ambiente externo. O simples fato de realizar os ensaios de interação fora do laboratório não implica necessariamente em que o usuário estará em uma condição mais próxima à condição real de uso. A condição real de uso de um

computador de mão não é única, há várias possíveis situações. Neste experimento foi definida uma delas, com o usuário sentado usando o computador de mão em um local público, bastante movimentado e com várias fontes de possível distração. Entretanto, o fato do avaliador estar presente aliado à utilização dos equipamentos para registrar a interação, fez com que o participante aumentasse a concentração na tarefa e não estivesse tão suscetível às interferências externas como era de se esperar. A pesquisa de novas técnicas que possibilitem ao usuário utilizar um computador de mão em situações reais, ou seja, em movimento, em vários ambientes diferentes, estando exposto às interferências externas e sem a presença do avaliador, ao mesmo tempo que permitam registrar a interação são importantes temas para a realização de trabalhos futuros. Algumas tentativas neste sentido podem ser citadas, como os trabalhos de Nielsen (2000) e Waterson (2002).

Finalmente as diferenças nos resultados apresentados no contexto da câmera de documentos mostraram que a situação de teste com o telefone apoiado sobre um tripé móvel, não solidário à câmera de documentos, influenciou o resultado dos testes. Esta constatação pode levar a duas conclusões.

A primeira conclusão é que estes resultados confirmam as suspeitas de Nyssonnen (2002) e Mizobuchi (2002) para quem a posição do usuário e a maneira como ele segura um computador de mão são fatores que afetam a interação com a interface e devem ser considerados. Os autores acreditam que o usuário móvel prefere segurar o equipamento nas mãos e ter liberdade para se movimentar, como eles fariam se estivessem em uma situação normal de

uso. Os resultados aqui obtidos sugerem que estas suspeitas são verdadeiras, uma vez que ter que operar o telefone sobre o tripé foi uma situação pouco confortável para o participante, que teve influência nos resultados da usabilidade medida.

A segunda conclusão é que a realização de tarefas adicionais paralelas podem interferir nos resultados da usabilidade medida e dos índices de desempenho das avaliações. Para tarefas explícitas, o experimento de Brewster (2002) já havia constatado uma queda na eficácia e um aumento na carga de trabalho quando foi solicitado aos participantes que andassem enquanto utilizavam um computador de mão. Entretanto, os resultados obtidos nesta pesquisa mostram que uma tarefa cognitiva concorrente, por menor que ela seja, como a preocupação em manter o tripé no foco da câmera de documentos, pode impactar negativamente sobre a usabilidade medida, fazendo aparecer problemas de usabilidade levemente “escondidos” por um contexto mais favorável. Estes dados sugerem novas abordagens nas avaliações de usabilidade para os computadores de mão. Por exemplo, a introdução de tarefas adicionais que exijam do usuário uma maior carga cognitiva, que não seja muito elevada mas que ao mesmo tempo seja capaz de favorecer o aparecimento de problemas de usabilidade que estariam na iminência de aparecer e que se manifestam somente nestas condições.

6.4 Recomendações para trabalhos futuros

Neste trabalho foram escolhidos três entre os possíveis contextos de avaliação que podem ser definidos para avaliar a usabilidade de uma aplicação WAP para telefone celular. A partir dos resultados aqui obtidos, para trabalhos futuros, sugere-se a definição de novos contextos de avaliação como os apresentados a seguir:

- Para este experimento foram escolhidas três abordagens representativas na definição dos contextos de avaliação para avaliar a usabilidade dos computadores de mão. Esta escolha se deu a partir da constatação do expressivo uso do emulador e da câmera de documentos como métodos para registrar a interação com os computadores de mão dentro do laboratório enquanto que a minicâmera sem fio foi a solução encontrada para registrar a interação na realização dos testes fora do laboratório. Estes contextos diferiam basicamente quanto aos seguintes aspectos: o equipamento utilizado para a navegação no portal (telefone celular ou emulador), o local de realização dos ensaios de interação (laboratório de usabilidade ou em campo) e os equipamentos utilizados para registrar a interação (computador, câmera de documentos ou câmera sem fio). As demais componentes dos contextos de avaliação foram exatamente as mesmas para as três abordagens. A partir dos resultados obtidos neste trabalho, sugere-se repetir o experimento utilizando somente o telefone celular e a câmera sem fio para realizar os testes dentro do laboratório e em campo. Ou seja, a diferença entre os dois contextos ficaria

reduzida a uma única variável que seria o local de realização dos testes. Desta forma seria possível obter novos dados sobre o impacto efetivo do local de realização dos testes nos resultados das avaliações de usabilidade dos computadores de mão.

- Há uma enorme quantidade de modelos de computadores de mão para os quais não existem emuladores cujas interfaces representem com fidelidade o equipamento e a interface real que se deseja avaliar. Portanto, na prática muitas vezes as avaliações de usabilidade vêm sendo conduzidas utilizando emuladores e telefones celulares com interfaces diferentes. Os resultados deste experimento indicaram que é possível obter bons resultados com o uso de emuladores desde que seja assegurada a similaridade entre as interfaces do emulador e do telefone celular, pois pequenas diferenças existentes podem impactar na usabilidade. Como nem sempre esta condição é possível de ser aplicada na prática, sugere-se um novo experimento para avaliar quais seriam os resultados da avaliação de usabilidade utilizando interfaces para o emulador e para o telefone celular completamente diferentes. Os resultados desse estudo permitiriam testar a hipótese de que a validade de uma avaliação de usabilidade utilizando um emulador só poderá ser garantida se as interfaces forem similares.

- Neste experimento foram definidos três contextos de avaliação controlados, nos quais não foi imposto ao participante a execução de nenhuma tarefa concorrente explícita seja dentro ou fora do laboratório. Os resultados

indicaram que, de uma maneira geral, para os contextos específicos aqui definidos não foi identificada uma diferença significativa em termos da usabilidade medida e dos índices de desempenho das avaliações. Entretanto, são inúmeras as possíveis situações de uso de um computador de mão, sendo comum a sua utilização enquanto o usuário executa atividades paralelas. Os resultados aqui obtidos indicaram que uma tarefa cognitiva concorrente, por menor que ela seja, pode ter um impacto negativo na usabilidade medida. A partir dessas constatações sugere-se um novo experimento que compare contextos de avaliação para os quais seja possível definir situações mais conturbadas. Seriam situações próximas à realidade nas quais o participante poderia estar executando tarefas concorrentes ou até mesmo estar sujeito a várias interrupções, enquanto executa a tarefa principal. Estas situações devem ser possíveis de ser realizadas tanto dentro quanto fora do laboratório. A definição destes contextos deve ser feita com cuidado pois eles devem se aproximar de possíveis situações reais de uso dos computadores de mão fora do laboratório que possam, ao mesmo tempo, ser reproduzidas dentro do laboratório, para que sejam compatíveis. Esse novo experimento permitiria testar a hipótese sugerida por este trabalho de que também não haveria diferença significativa entre os resultados das avaliações realizadas dentro e fora do laboratório nesses novos contextos de avaliação.

REFERÊNCIAS

ABNT. **Guia para utilização das normas sobre avaliação de qualidade de produto de software – ISO/IEC 9126 e ISO/IEC 14598**. ABNT-Associação Brasileira de Normas Técnicas. Curitiba, Maio, 1999.

ANATEL. PASTE/2000-Perspectivas para Ampliação e Modernização do Setor de Telecomunicações para o período 2000/2005. ANATEL-Agência Nacional de Telecomunicações. Brasília, 2000.

Ara, A. B., Musetti, A.V., Schneiderman, B. **Introdução à Estatística**. São Paulo: Instituto Mauá de Tecnologia, Edgard Blucher Ltda, 2003.

Arquer, I., Nogareda, C. **Estimación de la carga mental de trabajo: el método NASA TLX**. NTP-e 544, 2001.

Bastien, J.M.C., Scapin, D. **Evaluating a user interface with ergonomic criteria**. International Journal of Human-Computer Interaction, 7, 105-121, 1995.

Bergman, E. **Information Appliances and Beyond: interaction design for consumer products**. San Diego, CA: Academic Press, 2000.

Bevan, N. **Usability is Quality of use**. Proceedings of the 6th International Conference on Human Computer Interaction. Yokohama. July, 1995.

Bevan, N. **Common Industry format usability tests**. Proceedings of UPA'98. Scottsdale, Arizona. 29-June- 2July, 1999a.

Bevan, N. **Quality in use: meeting user needs for quality**. Journal of System and Software, 49 (1), 89-96, 1999b.

Bevan, N., Bogomolni, I. **Incorporating user quality requirements in the software development process**. 4th International Software & Internet Quality Week Conference-QWE 2000. Brussels, Belgium, 20-24 November, 2000.

Bevan, N. **International standards for HCI and usability**. International Journal of Human Computer Studies, 55(4), 533-552. October, 2001.

Bevan, N. **Cost-effective user-centred design based on ISO13407**. Tutorial notes. UPA'2002, Orlando, Florida,USA. July, 2002.

Bevan, N. **Setting Usability Performance Requirements**. Tutorial notes. CHI 2003, Fort-Lauderdale, Florida, USA. April, 2003.

Bocker, M. , Suwita, A. **Evaluating the Siemens C10 Mobile Phone: Going Beyond “Quick and Dirty” Usability Testing**. Proceedings of the 17th International Symposium on Human Factors in Telecommunication. Copenhagen, Denmark. May, 1999.

Brag, Wilhelm, Wentrup, Robert. **Developing Wapps-wireless applications in the context of systems development.** Master thesis – Department of Informatics, School of Economics and Commercial Law, Göteborg University, 2001.

Brewster, S. **Overcoming the lack of screen space on mobile computers.** Personal and Ubiquitous Computing 6 (3) May, 2002.

Brooke, J. **SUS – A quick and dirty usability scale.** In: Jordan, P.W. et al., Usability Evaluation in Industry, pp. 189-94. London, UK: Taylor & Francis, 1996.

Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G., Pazzani, M. **Improving Mobile Internet usability.** Proceedings of the 10th International World wide web Conference. Hong Kong, China. May-1-5, 2001.

Caulton, D. A. **Relaxing the homogeneity assumption in usability testing.** Behaviour & Information Technology, vol. 20, n^o 1, 1-7, 2001.

Chincholle, D., Goldstein, M. **The Finger-Joint-Gesture Wearable Keypad,** INTERACT'99 - Workshop on Mobile Devices, Edinburgh, Scotland, August 1999.

Chincholle, D., Eriksson, M. **Designing Highly Usable Wireless Services for Next Generation Devices and Networks.** Tutorial notes. UPA'2002, Orlando, Florida,USA. July, 2002a.

Chincholle, D., Goldstein, M., Nyberg, M., Eriksson, M. **Lost or found? A usability evaluation of a mobile navigation and location-based service.** Proceedings of Mobile HCI 2002 - Fourth International Symposium on Human Computer Interaction with Mobile Devices. Pisa, Italy, September 2002b.

Chittaro, L., Cin, P. **Evaluating interface design choices on WAP phones: single-choice list selection and navigation among cards.** Proceedings IHM-HCI Mobile HCI 2001.

CIF-Common Industry Format for Usability Test Reports. American National Standard for Information Technology. USA: ANSI, November, 2001.

Cooper, 1995. **Performance Measurement Handbook.** National Physical Laboratory, Teddington, Middlesex, U.K. 1995.

Cybis, W. **Ergonomia de Interfaces Humano-Computador.** Florianópolis: 2002.

Davies, M. **A usability test by 3G LAB: comparison of two cameras phones: Nokia 7650 vs. Sony-Ericsson T68i.** 3G LAB. Cambridge, UK. September, 30th, 2002.

Desurvire, H. W. **Faster, Cheaper!! Are usability inspection methods as effective as empirical testing?** In: Nielsen, J., Mack, R.L. **Usability Inspection Methods.** USA: John Wiley & Sons, 1994. p. 173-202.

Digitwireless. Disponível em: <http://www.digitwireless.com>. Acesso em: março de 2003.

Dix, A. et al. **Exploiting context in HCI design for mobile systems**. In C. Johnson (Ed.), Proceedings of First Workshop on HCI for Mobile Devices, GIST technical Report G98-1, department of Computer Science, University of Glasgow, Scotland, 12-17. 1998.

Duda, S., Schiessl, M. Hess, J. M. **Mobile Usability**. Eye Square & Mobile Economy. Berlin, Germany. 12 April, 2001.

Dumas, J.S., Redish, J.C. **A Practical Guide to Usability Testing**. England: Intellect, 1999.

Ericsson, T. Chincholle, D., Goldstein, M. **Both the celular phone and the service impact WAP usability**. Proceedings of IHM-HCI 2001, Volume II, HCI in Practice, by J. Vanderdonckt, A. Blanford and A. Derycke (eds). Lille, France, 10-14 September, 2001.

Fossil. Disponível em: <http://www.fossil.com>. Acesso em: março de 2003.

Gaffney, G. **The InfoDesign Toolkit**. Australia: Information & Desing, 2000.

Gil, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 1991.

GSA. **Survey of Mobile Portal Services, Q42002**. Global Mobile Suppliers Association, UK, 2002.

GSM. Disponível em: <http://www.gsmworld.com/news/statistics>. Acesso em: janeiro de 2003.

Hackos, J.T., Redish, J. C. **User and Task Analysis for Interface Design**. USA: John Wiley & Sons, 1998.

Handspring. Disponível em: <http://www.handspring.com>. Acesso em: janeiro de 2003.

Hansell, S. **A Microsoft Watch Will Provide Much More Than Time**. New York Times, Late Edition - Final , Section C , Page 4 , Column 1. New York, January, 9th 2003

Helyar, V. **Usability issues and user perceptions of a 1st generation WAP service**. Serco Usability Services, London. January, 2001.

Helyar, V. **Assessing user requirements for the mobile internet**. Appliance Design, issue3: Mobile Internet. Bill Sharpe ed. The Appliance Design Network, UK: September, 2002

Hewett et al. **ACM-SIGCHI Curricula for Human-Computer Interaction**. 1996.

Hiltunen, M., Laukka, M., Luomala, J. **Mobile User Experience**. Finland: Edita Publishing Inc., 2002.

Holmquist, L. E. Future Applications Lab - Viktoria Institute, Suécia. Disponível em: <http://www.viktoria.se/~leh>. Acesso em: agosto de 2002.

Hughes, M. **Rigor in Usability Testing**. Technical Communication, fourth quarter, 1999.

Ipiña, D. L., Mendonça, P.R., Hopper, A. **TRIP: A low-cost vision-based location system for Ubiquitous computing**. Personal and Ubiquitous Computing 6 (3) May, 2002.

IDC. **IDC Brasil conclui estudo do mercado nacional de telefonia celular**. Press release da IDC Brasil referente ao estudo Brazil Wireless Market Sizing. Abril, 2002.

ISO/IEC 9126-1 – Software Product Quality – Part 1: Quality Model, 2001.

ISO 9241-11 – Ergonomic Requirements for office work with visual display terminals (VDT)s – Part 11: Guidance on usability, 1998.

ISO 13407 Human-Centred Design process for Interactive Systems, 1999.

ISO/IEC 14598-1 - Software product evaluation - Part 1: General overview, 1999.

ISO/TR 18529 - Human-centred lifecycle process descriptions, 2000.

ITU Internet Reports 2002: **Internet for a mobile generation-executive summary**. International Telecommunication Union, September, 2002.

Jeffries, R., Miller, J., Wharton, C., Uyeda, K. **User interface evaluation in the real world: a comparison of four techniques**. Proceedings of CHI'91, New Orleans, April 28-May 3, 1991.

Johnson, P. **Usability and Mobility: interactions on the move**. First Workshop on Human-Computer Interaction with Mobile Devices. Glasgow, UK, 1998.

Jokela, T., Iivari, N., Matero, J., Karukka, M. **The standard of user-centered design and the standard definition of usability: analysing ISO 13407 against ISO 9241-11**. Proceedings of CLIHIC 2003, Rio de Janeiro, August 17-20, 2003.

Jordan, P.W. **An Introduction to Usability**. London, UK: Taylor & Francis, 1998.

Kaasinen, E., Aaltonen, M., Kolari, J., Melakoski, S., Laakko, T. **Two approaches to bringing Internet services to WAP devices**. Proceedings WWW9-9th International World Wide Web Conference, Amsterdam, May 15 - 19, 2000.

Karat, C. **A comparison of user interface evaluation methods**. In: Nielsen, J., Mack, R.L. **Usability Inspection Methods**. USA: John Wiley & Sons, 1994. p. 203-233.

Keinonen, T. **Usability of Artifacts**. In One dimensional usability – influence of usability on consumers product preference. Helsinki: UIAH publication A21, 1998. Cap. 2.

Kittytech. Disponível em: <http://www.kittytech.com>. Acesso em: janeiro de 2003.

Lavery, D., Cockton, G., Atkinson, M.P. **Comparison of evaluation methods using structured usability problem reports.** Behaviour & Information Technology, 16, 4/5, 246-266. 1997.

Little Springs Design Disponível em: <http://www.littlespringsdesign.com/design/textinput.html>. Acesso em: janeiro de 2003.

MacKenzie, I.S., Soukoreff, R.W. **Text entry for mobile computing: Models and methods, theory and practice.** Human-Computer Interaction, 17, 147-198. 2002.

Macleod, M., Bowden, R., Bevan, N., Curson, I. **The MUSiC performance measurement method.** Behaviour & Information Technology, vol. 16, nº 4/5, 279-293. 1997.

Maguire, M. **Context of use within usability activities.** International Journal of Human-Computer Studies, 55(4), 453-483. 2001.

Mayhew, D.J. **The Usability Engineering Lifecycle.** San Francisco, CA: Morgan Kaufmann, 1999.

McIntyre, L. Lee. **Remote Usability Testing for a Wireless Device: an unorthodox protocol.** Poster presented at UPA'2002, Orlando, Florida, USA. July, 2002.

Microsoft. Disponível em: <http://www.microsoft.com/resources/spot/product.mspc>. Acesso em: janeiro de 2003.

Mizobuchi, S., Mori, K., Ren, X., Michiaki, Y. **An empirical study of the minimum required size and the minimum number of targets for pen input on the small display.** Proceedings of Mobile HCI 2002 - Fourth International Symposium on Human Computer Interaction with Mobile Devices. Pisa, Italy, September 2002.

Mobinet #5. **Mobinet Index #5.** A.T. Kearney, University of Cambridge-The Judge Institute of Management Studies.UK: August, 2002.

Mohageg, M. F., Wagner, A. **Design consideration for Information Appliances.** In: Bergman, Eric. Information Appliances and Beyond: interaction design for consumer products. San Diego, CA: Academic Press, 2000.

NASA TLX - NASA Task Load Index. Human Performance Research Group, NASA Ames Research Center.Moffett Field, California. 1986.

NBR 9241-11. **Requisitos ergonômicos para trabalho em escritórios com computadores Parte 11 – Orientações sobre Usabilidade.** (Norma equivalente à ISO 9241-11: 1998). ABNT, 2002.

Nielsen, J. **Usability Engineering.** San Francisco, CA: Morgan Kaufmann, 1993a.

Nielsen, J. **Noncommand user interfaces.** Communications of the ACM 36, 4, 83-89, April, 1993b.

Nielsen, J. Landauer, T.K. **A mathematical model of the finding of usability problems.** INTERCHI'93. Amsterdam, the Netherlands, 24-29 April, 1993c.

Nielsen, J., Mack, R.L. **Usability Inspection Methods.** USA: John Wiley & Sons, 1994.

Nielsen, J., Ramsay, M. **WAP Usability report.** Nielsen Norman Group, December, 2000.

Nielsen, J. **Why you only need to test with five users.** Alertbox March, 2000. Disponível em: <http://www.useit.com/alertbox/20000319.html>. Acesso em: maio 2002.

Nielsen, J. **Paper prototyping: a how-to training video.** Nielsen Norman Group, 2003a.

Nielsen, J. **230 Tips and Tricks for a Better Usability Test.** Nielsen Norman Group, 2003b.

Norman, D. A. **The Psychology of Everyday Things.** USA: Basic Books, 1988.

Norman, D. A. **The Invisible Computer.** Cambridge, MA: The MIT Press, 1998.

Nokia. Disponível em: <http://www.nokia.com>. Acesso em: janeiro de 2003.

NTTDoCoMo. Disponível em: <http://www.nttdocomo.com> e <http://www.nttdocomo.co.jp>. Acesso em: janeiro de 2003.

Nyberg, M., Goldstein, M., Leung, Y. **Visualising data using the ActionMapper: a proposed interactive event logger for user interface evaluation.** International Conference on Information Visualisation, IV 2001. London, England, UK, 25-27 July 2001.

Nyyssönen, T., Roto, V. Kaikkonen, A. **Mini-camera for Usability Tests and demonstrations.** Demonstrated at Mobile HCI 2002 - Fourth International Symposium on Human Computer Interaction with Mobile Devices. Pisa, Italy, September 2002.

OMA. The official web site for The Open Mobile Alliance Ltd. Disponível em <http://www.openmobilealliance.org/>. Acesso em: fevereiro de 2003.

Öörni, K. **What do we know about usability evaluation? A critical view.** DIGILIB 2003. 2003 Conference on Users in the Electronic Information Environments. Espoo, Finland, September, 8-9, 2003.

Öquist, G., Goldstein, M. **Towards an improved readability on mobile devices: evaluating adaptative rapid serial visual presentation.** Proceedings of Mobile HCI 2002 - Fourth International Symposium on Human Computer Interaction with Mobile Devices. Pisa, Italy, September 2002.

Pearrow, Mark. **The Wireless Web Usability Handbook.** Hingham, MA: Charles River Media, 2002.

Perfetti, C. Landesman, L. **Eight is not enough**. Disponível em: [http://world.std.com/~uieweb/Articles/eight is not enough.htm](http://world.std.com/~uieweb/Articles/eight%20is%20not%20enough.htm). Acesso em: março, 2003.

Pers,J., Hove, W., Theuws, J. Noldus, L. **A new wireless camera for usability evaluation of mobile devices**. Presented at Mobile HCI 2002 - Fourth International Symposium on Human Computer Interaction with Mobile Devices. Pisa, Italy, September 2002.

Petrie, H., Furner, S., Strothotte, T. **Design lifecycles and wearable computers for users with disabilities**. First Workshop on Human-Computer Interaction with Mobile Devices. Glasgow, UK, 1998.

QUIS-Questionnaire for User Interaction Satisfaction. Disponível em: <http://www.cs.umd.edu/hcil/quis/>. Acesso em: fevereiro de 2003.

Rao, D.O. **A study of input devices used on personal digital assistants (PDAs)**. Serco Usability Services, London, 2000.

Rischpater, R. **Wireless Web Development**. Berkeley, CA: Apress, 2000.

Rocha, H. V., Baranauskas, M.C. **Design e Avaliação de Interfaces Humano-Computador**. Campinas: NIED, 2003.

Rubin, J. **Handbook of Usability Testing: how to plan, design, and conduct effective tests**. USA: John Wiley & Sons, Inc., 1994.

Savage, P. **User interface evaluation in an iterative design process: a comparison of three techniques**. CHI'96 short paper. Conference on Human Factors in Computing Systems. Vancouver, Canada, April, 13-18, 1996.

Scapin, D., Berns, T. **Usability evaluation methods**. Behaviour & Information Technology, vol.16, (4/5), 185-187, 1997.

Sears, A., Zha, Y. **Data entry for mobile devices using soft keyboards: understanding the effects of keyboard size and user tasks**. International Journal of Human-Computer Interaction, 16 (2), 163-184, 2003.

Sears, A. **Heuristic Walkthroughs: finding the problems without the noise**. International Journal of Human-Computer Interaction, 9 (3), 213-234, 1997.

Senseboard. Disponível em: <http://www.senseboard.com>. Acesso em: janeiro de 2003.

Shackel, B. **Usability – context, framework, design and evaluation**. In Shackel, B. and Richardson, S. (eds.). Human Factors for Informatics Usability. Cambridge University Press, Cambridge, 21-38, 1991.

Silfverberg, M. **The one-row keyboard - a case study in mobile text input**. In: Lindholm, C., Keinonen, T., Kiljander, H. Mobile Usability – How Nokia changed the face of the mobile phone. New York, NY, USA: Mc Graw Hill, 2003.

Sony Ericsson. Disponível em: <http://www.sonyericsson.com>. Acesso em: março de 2003.

Spool, J., Schroeder, W. **Testing Web sites: five users is nowhere near enough.** CHI 2001, Extended Abstracts. Seattle, March 31- April 5, 2001.

Spyridakis, J.H. **Conducting Research in Technical Communication: The Application of True Experimental Designs.** Technical Communication, 39, 4, 607-624, 1992.

SUMI-Software Usability Measurement Inventory. Disponível em: <http://www.ucc.ie/hfrq/questionnaires/sumi/>. Acesso em: fevereiro de 2003.

Sutcliffe, A. Doubleday, A. Ryan, M. Springett, M. **A comparison of usability techniques for evaluation design.** Proceedings of Designing Interactive Systems: Process, Practices, Methods and Techniques. ACM Press. The Netherlands: August, 1997.

SUS-System Usability Scale. Disponível em: <http://www.usability.serco.com/trump/methods/satisfaction.htm>. Acesso em: fevereiro de 2003.

Svanteson, S. **Particular Challenges in Designing for Baby-Faces.** Master thesis written for the Department of Media Technology and Graphics Arts, Royal Institute of Technology in Stockholm and Razorfish. Spring, 2000.

Taurion, C. **Internet Móvel: tecnologias, aplicações e modelos.** Rio de Janeiro: Campus, 2002.

Taurion, C. **Cenários das aplicações móveis para os próximos anos.** Mobile Corporate Forum. Curitiba, agosto, 2003.

T9. AOL Mobile. Disponível em: <http://www.t9.com>. Acesso em: janeiro de 2003.

Usability Systems Inc. Disponível em: <http://www.usabilitysystems.com>. Acesso em: março de 2002.

Usablemobile. Lista de discussão sobre usabilidade para os computadores de mão. Disponível em: <http://www.usablemobile.com>. Acesso em: janeiro de 2003.

Väänänen, K., Vainio-Mattila, Ruuska, Satu. **Designing Mobile Phones and Communicators for Consumers' Needs at Nokia.** In: Bergman, Eric. Information Appliances and Beyond: interaction design for consumer products. San Diego, CA: Academic Press, 2000.

Valdehita, S.R., Ramiro, E.D. **La medida de la carga mental de trabajo II: Procedimientos subjetivos.** Boletín Factores Humanos, 21, Diciembre, 1999.

Valente, A. C. **Perspectivas Futuras dos Serviços Móveis-ANATEL.** III Seminário de Qualidade do Serviço Móvel Celular e Pessoal. Brasília, 20 e 21 de novembro de 2002.

Virzi, R. A. **Refining the test phase of usability evaluation: how many subjects is enough?** Human Factors 34 (4), 1992.

VKB. Disponível em: <http://www.vkb.co.il>. Acesso em: janeiro de 2003.

Vyas, D. **Investigating usability issues in WAP applications.** MSc Dissertation for Advanced MSc in Distributed Interactive Systems, Lancaster University, September, 2002.

Waterson, S. , Landay, J.A., Matthews, T. **In the lab and out in the wild: remote Web usability testing for mobile devices.** Extended Abstracts of ACM CHI 2002 Conference on Human Factors in Computing Systems. Minneapolis, MN. April, 2002.

Weiser, M. **The Computer for the Twenty-First Century.** Scientific American, pp. 94-10, September 1991

Weiss, S., Kevil, D., Martin, R. **Wireless phone usability findings.** Usable Products, New York, January, 2001.

Weiss, S. **Handheld Usability.** England: John Wiley & Sons Ltd, 2002.

Winter, E.M.W. **Introdução à Bioestatística.** Departamento de Estatística, Universidade Federal do Paraná. 2002.

Woolrych, A., Cockton, G. **Why and when five test users aren't enough.** Proceedings of IHM-HCI 2001 Conference. Toulouse, France, 2001.

APÊNDICE A

Roteiro com as informações fornecidas ao participante

Muito obrigado pela sua presença. Você está aqui hoje para me ajudar a avaliar o uso de algumas aplicações de Internet móvel para telefones celulares. Através do telefone celular e da tecnologia WAP é possível ter acesso a uma série de informações e serviços como e-mail, notícias, lazer, esportes, transações bancárias, etc. Eu vou pedir que você utilize alguns destes serviços e depois você vai me dar a sua opinião sobre o que você achou dessa experiência. O nosso objetivo é colher a sua opinião em relação à usabilidade destas aplicações.

Eu sei que esta não é uma situação usual para você mas por favor tente agir da maneira mais natural possível, como você agiria se estivesse sozinho.

Eu lhe darei as tarefas, uma de cada vez, e você vai executá-las normalmente, como você faria se estivesse realmente naquela situação. Quando você considerar que completou a tarefa, peço que você fale em voz alta a informação que você encontrou.

Você pode fazer perguntas, entretanto nem sempre eu poderei respondê-las, pois nosso objetivo é simular o uso deste sistema supondo que você está sozinho, por isso você trabalhará sozinho, sem o meu auxílio. Não se preocupe com resultados, não há respostas certas nem erradas, lembre-se que você está aqui para me ajudar a avaliar a usabilidade desta aplicação.

Estaremos filmando as imagens da tela do computador (ou da tela e teclado do telefone) e também gravaremos o áudio desta sessão.

Esta sessão durará cerca de uma hora e meia e se você precisar interrompê-la a qualquer momento, é só me avisar.

APÊNDICE B

Questionário 'Perfil do Usuário'

Idade

- até 19 anos
- de 20 a 25 anos
- de 26 a 30 anos
- de 31 a 35 anos
- de 36 a 40 anos
- mais de 41 anos

Sexo

- Masculino
- Feminino

Qual seu grau de instrução?

- Até 1º grau incompleto
- 1º grau completo
- 2º grau incompleto
- 2º grau completo
- Superior incompleto
- Superior completo ou mais

Qual sua formação? _____

Qual sua ocupação?

- Funcionário de empresa privada
- Funcionário público
- Profissional liberal
- Autônomo
- Empresário
- Estudante
- Aposentado
- Desempregado

Você utiliza computador?

- Sim
- Não

Há quanto tempo você acessa a Internet?

- Nunca usei
- Menos de um ano
- De 1 a 3 anos
- De 3 a 5 anos
- Mais de 5 anos

Com que frequência você acessa a Internet?

- Nunca
- 1 vez por mês
- 1 vez por semana
- 1 vez por dia
- Mais de 1 vez por dia

Qual o seu tempo, em média, por acesso à Internet?

- Eu não acesso a Internet
- Menos de 1 hora
- De 1 a 3 horas
- De 3 a 5 horas
- De 5 a 7 horas
- De 7 a 9 horas
- Mais de 9 horas

Há quanto tempo você usa e-mail?

- Nunca usei
- Menos de um ano
- De 1 a 3 anos
- De 3 a 5 anos
- Mais de 5 anos

Você tem telefone celular?

- Sim Qual a sua operadora ? Vivo TIM Outra
- Não

Há quanto tempo você é usuário de telefone celular?

- Menos de um ano
- De 1 a 3 anos
- De 3 a 5 anos
- Mais de 5 anos

O seu aparelho de telefone celular tem acesso a WAP?

- Sim Não
- Não sei

Com que frequência você lê mensagens de texto no seu telefone celular?

- Nunca
- 1 vez por mês
- 1 vez por semana
- 1 vez por dia
- Mais de 1 vez por dia

Com que frequência você envia mensagens de texto do seu telefone celular?

- Nunca
- 1 vez por mês
- 1 vez por semana
- 1 vez por dia
- Mais de 1 vez por dia

Com que frequência você utiliza seu telefone celular para acessar aplicações**WAP?**

- Nunca
- 1 vez por mês
- 1 vez por semana
- 1 vez por dia
- Mais de 1 vez por dia

Você já usou o telefone celular LG Sensations, modelo que será utilizado neste experimento?

- Sim Não

APÊNDICE D

Questionário 'Avaliação Geral sobre WAP'

	Discordo Completamente	Discordo	Neutro	Concordo	Concordo completamente
1. A navegação em um telefone celular não é um problema quando você se acostuma com o aparelho.					
2. Um tempo de conexão rápido é importante para que eu use o WAP.					
3. O tamanho da tela do telefone celular é um fator crítico que pode influenciar a minha escolha entre usar ou não o WAP.					
4. O tamanho do teclado do telefone celular é um fator crítico que pode influenciar a minha escolha entre usar ou não o WAP.					
5. A qualidade do conteúdo é um fator crítico que pode influenciar a minha escolha entre usar ou não o WAP.					
6. O custo de acesso é um fator crítico que pode influenciar a minha escolha entre usar ou não o WAP.					
7. Eu já tinha conhecimento sobre todos os diversos serviços que estão disponíveis no WAP.					
8. As novas tecnologias permitirão que eu receba em meu telefone serviços personalizados e baseados na minha localização geográfica. Isto me encoraja a usar os serviços WAP no futuro.					
9. Logo será possível me localizar onde quer que eu esteja, a partir do meu telefone celular. Isto pode ser considerado intrusivo e fazer com que eu deixe de usar os serviços WAP.					

ANEXO A

Questionário de Satisfação SUS – System Usability Scale

	Discordo completamente	Discordo	Neutro	Concordo	Concordo completamente
1. Eu penso que gostaria de usar este sistema freqüentemente.					
2. Achei este sistema desnecessariamente complexo.					
3. Achei que foi fácil usar este sistema.					
4. Eu penso que precisaria de ajuda para poder usar este sistema.					
5. Achei que as várias funções deste sistema estavam bem integradas.					
6. Achei que havia muita inconsistência neste sistema.					
7. Eu imagino que a maioria das pessoas aprenderia a usar este sistema rapidamente.					
8. Achei este sistema muito incômodo de usar.					
9. Eu me senti muito seguro(a) utilizando este sistema.					
10. Eu precisei aprender muitas coisas antes de utilizar este sistema.					