

Instituto Superior Técnico

Interfaces Pessoa Máquina

Análise de resultados de testes com utilizadores

Grupo 1:

- António Vieira 86387
- Jorge Pacheco 86457
- Pedro Campos 86494

Preparação

Os testes destinam-se a ser realizados utilizando os computadores dos elementos do grupo, onde o protótipo se encontra melhor otimizado, espera-se que, para além destes computadores, seja apenas necessário um aparelho capaz de cronometrar tempo e os questionários pré e pós-teste.

Introdução

Este guião destina-se a descrever o procedimento do teste do dispositivo *iRave*, assim como os cenários de uso a que os utilizadores de teste estarão sujeitos. Este teste destina-se a avaliar os níveis de eficiência e eficácia do dispositivo desenvolvido. Aos participantes serão pedidas apenas informações gerais como idade, nível de experiência com dispositivos, nível de capacidade de interiorização, entre outros.

Avaliação

Antes da execução dos testes, para além de serem indicados os cenários de teste, serão explicados aos utilizadores algumas instruções relativas à navegação do dispositivo (por exemplo, o *scroll* deve ser feito com o rato, o botão da esquerda retorna para trás e o da direita liga/desliga o dispositivo).

Os cenários de teste estão agrupados em três categorias, associadas às três funcionalidades principais do protótipo: adição e localização de pessoas, ordem de pedidos e verificação de notificações.

Os testes a serem executados com os utilizadores serão os seguintes:

1. Adicionar 3 amigos, confirmar que foram adicionados e encontrar um caminho para um deles no mapa.

Tendo em conta que a localização no dispositivo se encontra desligada por omissão, espera-se que esta seja a tarefa que requeira mais tempo para ser executada, dado que não pode ser completada utilizando uma só aplicação e o utilizador necessita de encontrar as definições da localização. Espera-se que o utilizador necessite de 1 minuto para realizar esta tarefa.

É de notar que neste teste será dada mais importância à eficiência observada, dado que este é o teste que introduz maior margem de erros possíveis.

2. Comprar três cervejas e uma sangria, e pedir um lugar na fila do WC.

Este é outro teste que requer mais que uma aplicação, contudo irá requerer menos tempo por parte do utilizador, dada a natureza das aplicações em questão. Espera-se que o utilizador necessite de 30 segundos para realizar esta tarefa.

3. Seguir dois artistas à escolha, verificar as suas notificações, deixar de seguir um e seguir mais um, e verificar a nova notificação.

Este é o único teste que não requer sair da aplicação principal do objetivo pedido, portanto espera-se que não seja o teste que requeira mais tempo. Espera-se que o utilizador necessite de 50 segundos para realizar esta tarefa.

Além disto, também será tido em conta o *feedback* dos utilizadores, no que toca à satisfação do protótipo desenvolvido, com um questionário pós-teste, constituído por 10 afirmações às quais o utilizador pode concordar ou discordar.

Balanço

Para além dos testes, o *feedback* dos utilizadores será obtido utilizando um SUS (*System Usability Scale*), que será composto por 10 afirmações relativas ao protótipo desenvolvido, às quais o utilizador pode avaliar numa escala de 1 (caso discorde totalmente) a 5 (caso concorde totalmente). As afirmações são:

1. Usaria frequentemente o *iRave* em festivais.
2. Penso que o sistema é desnecessariamente complexo.
3. Acho que o sistema é fácil de usar.
4. Penso que necessitaria de ajuda técnica para utilizar o sistema.
5. Achei que as funcionalidades estavam bem integradas no sistema.
6. Achei o *iRave* inconsistente.
7. Imagino que a maioria das pessoas aprenderia a usar o *iRave* rapidamente.
8. Achei o sistema pouco apelativo.
9. Senti-me confiante a usar o *iRave*.
10. Precisei de aprender muitas coisas antes de utilizar o sistema.

Caracterização dos utilizadores

A amostra de população que participou na avaliação consiste num grupo de 15 pessoas, cujas idades variavam entre os 18 e 25 anos, e das quais 86,7% eram do sexo masculino. Ademais, nesta amostra, a vasta maioria das pessoas frequenta 1 a 2 festivais por ano e possui um alto nível de experiência com dispositivos eletrónicos.

Resultados obtidos

Notou-se, em primeiro lugar, que as médias de tempo de realização para cada teste descrito foram inferiores, em média, por volta de 9,5 segundos, possivelmente devido ao elevado nível de experiência em dispositivos eletrónicos da amostra. Contudo, no primeiro e terceiro teste, os tempos medidos demonstravam uma variação de tempo mais acentuada (demonstrado pelos seus desvios padrão), enquanto que no segundo teste, os tempos individuais medidos aproximavam-se mais da média calculada.

Em suma, as médias de tempo, desvios padrão e intervalos de confiança para cada tarefa são as seguintes:

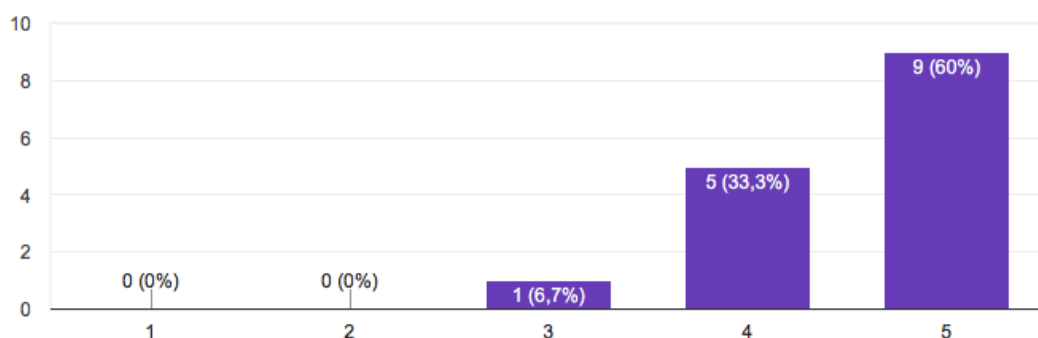
	Média de tempos	Desvio padrão	Intervalo de confiança
Tarefa 1	41,2 segundos	9,62 segundos	[36,6 ; 45,8]
Tarefa 2	19,0 segundos	3,03 segundos	[17,5 ; 20,5]
Tarefa 3	31,1 segundos	8,25 segundos	[27,1 ; 35,1]

É de notar que, para calcular os intervalos de confiança, foi escolhido um nível de confiança de 95%.

O *feedback* obtido dos mesmos utilizadores através do questionário pós-teste revelou ainda que grande parte dos mesmos achou o protótipo desenvolvido relativamente simples de utilizar e consistente, contudo, também foram registados casos em que os utilizadores achassem que necessitariam de ajuda técnica (possivelmente através de um manual de instruções, por exemplo). Também foram registados comentários em relação ao menu principal, que deveria ser extensível horizontalmente e não verticalmente, dada a natureza dos *smartwatches*, ademais, um atalho para ligar a localização poderia ser implementado, quando o dispositivo relembra que a localização se encontra desligada, o que reduziria, ainda mais, o tempo de execução da primeira tarefa.

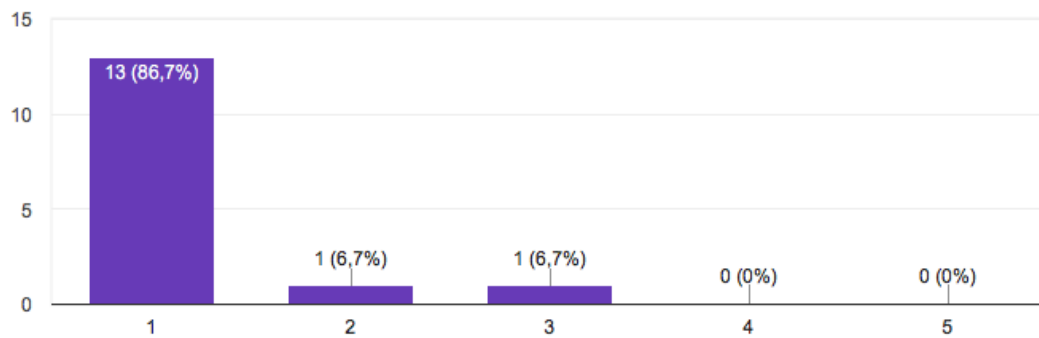
Acho que o sistema é fácil de usar.

15 respostas



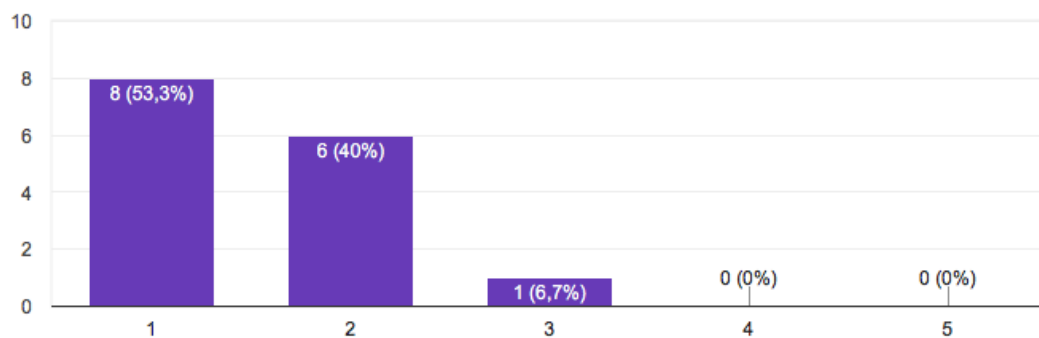
Penso que necessitaria ajuda técnica para utilizar o sistema.

15 respostas



Achei o iRave inconsistente.

15 respostas



Em conclusão, face ao *feedback* positivo dos utilizadores, poder-se-á dizer que a eficiência e eficácia do protótipo desenvolvido é relativamente alta.