

PhD Course Lecture

Usability measurement

Howell Istance
De Montfort University

Overview of my bit...

- Overview of usability measurement..
- Measuring Subjective attributes
- Applying this to a comparison of eye-based pointing and head-based pointing

Lecture 1: Overview

- What are we trying to measure and why?
- History
- Metrics and Impact Analysis
- Fitts Law and pointing device assessment
- Using eye gaze to assess usability

Old days....

- Early-Mid 1980's
- Pre-dominance of IBM PC
- Large numbers of administrative workers using WP and data entry programs
- Character-based interfaces (rather than graphical UI's)
- Command languages used frequently
- "User-friendly" – sales term – meant usually "has menus"

Visicalc

120 (U) +H20#12 19

	A	G	H	I
HOME BUDGET, 1979				
MONTH		NOV	DEC	TOTAL
SALARY		2500.00	2500.00	30000.00
OTHER				

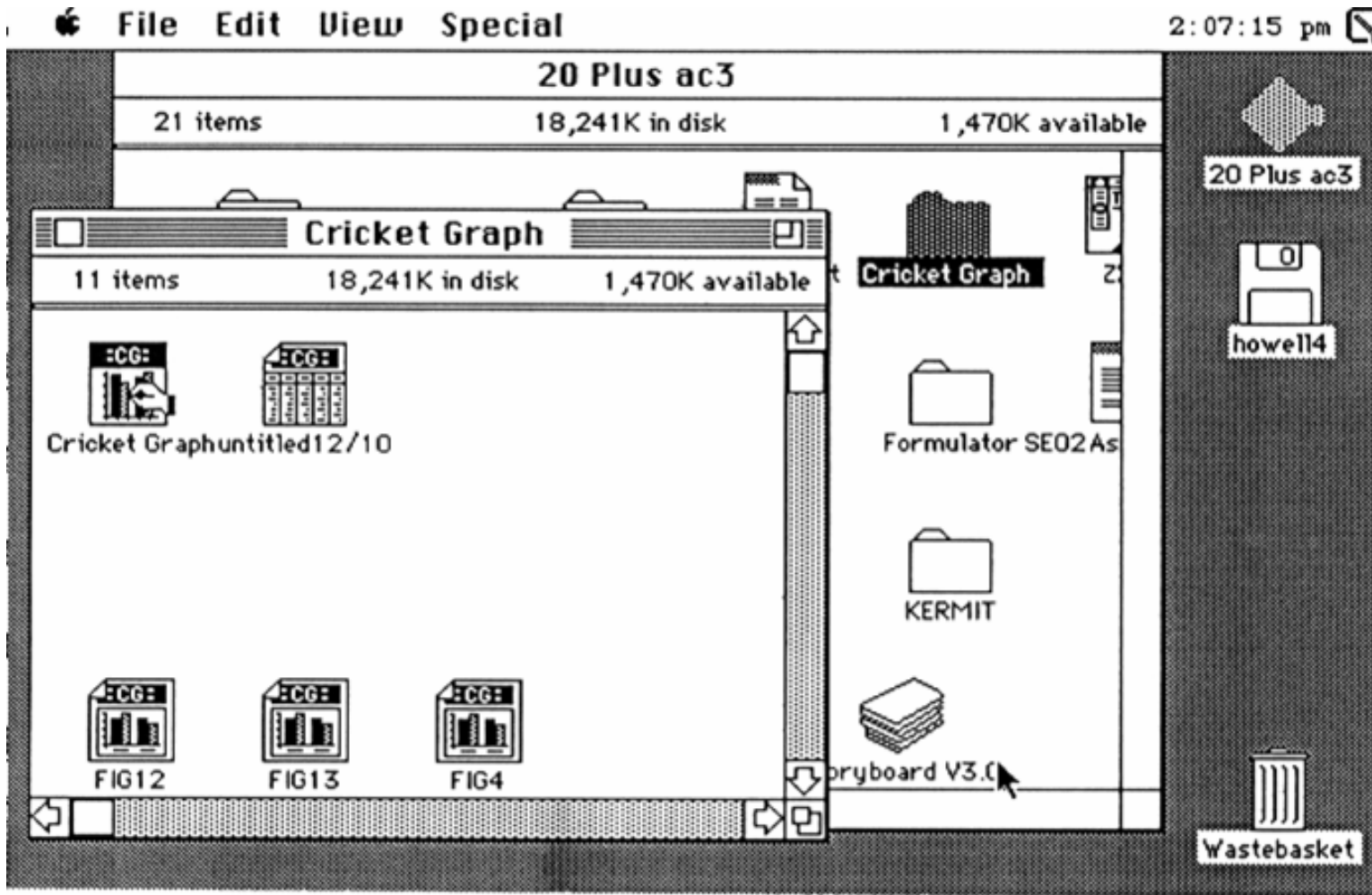
INCOME		2500.00	2500.00	30000.00
FOOD		400.00	400.00	4800.00
RENT		350.00	350.00	4200.00
HEAT		110.00	120.00	575.00
REC.		100.00	100.00	1200.00
TAXES		1000.00	1000.00	12000.00
ENTERTAIN		100.00	100.00	1200.00
MISC		100.00	100.00	1200.00
CAR		300.00	300.00	3600.00

EXPENSES		2460.00	2470.00	28775.00
REMAINDER		40.00	30.00	1225.00
SAVINGS		30.00	30.00	350.00

Advent of the MacIntosh

- Competed strongly with the PC in spite of being different/poorer (small mono-chrome screen, 3.5" disk, proprietary software, slow disk access)
- Graphical metaphor-based UI
- Emphasis on consistency across applications (MacWrite, MacPaint, MacDraw, MacProject...etc)

MacIntosh Desktop



Usability and Usability Engineering..

- Part of an attempt to put HCI into the software development lifecycle (as opposed to an optional nicety)
- Elevate usability to an aspect of the system that can be engineered to meet a given specification
- Important players John Whiteside, Karen Holtzblatt (DEC labs and John Bennett (IBM)

Measuring usability

- To test whether an engineering specification has been met
 - Basis for usability engineering
- To quantify the extent of a usability problem
 - Usually on an ordinal scale ('major', 'moderate', 'minor')

Usability Engineering

- aim is to identify for public discussion
 - what counts as success
 - how to measure success
- top level attributes of usability are related to
 - learnability (or ease of learning)
 - throughput (or ease and efficiency of use after learning)
 - user satisfaction
 - Or any set of attributes considered important by project personnel

Definitions (ISO-9241 part 11)

- Usability
 - The degree to which specific users can achieve specified goals in a particular environment with effectiveness, efficiency and satisfaction
- Effectiveness
 - measures of of the accuracy and completeness of the goals achieved
- Efficiency
 - measures of the resources (e.g time money, effort) used to achieve goals
- Satisfaction
 - measures of the physical comfort and subject acceptability of the product to its users

Objectives

- Give definition of usability through metrics
- Set planned levels for usability attributes
- Incorporate user-derived feedback into design process and revise design
- Continue until usability levels are met

Measurable quantities

- Time to complete a task
- Percent of task completed
- percentage of task completed per unit time
- ratio of successes to failures
- time spent in errors
- number of errors made
- frequency of help or documentation use
- duration of help or documentation use
- percentage of favourable/unfavourable user comments
- number of repetitions or failed commands
- number of good and features recalled by user
- number of commands not used
- number of times users need to work around a problem

Components of a usability specification

- statement of the usability attribute
- statement of how it will be measured
- statement of criteria that will represent attainment of the specification
- statement of the subset of users to which the specification applies
- statement of pre-conditions of measurement (e.g period of training)

Example of a usability specification

- In the context of a library information system e.g.

Efficiency of first time use measured by retrieval time of specified information

'90 % of student users without any prior training should be able to retrieve the shelf number of a book, given the author and the title within 10 times the length of time an expert user would take to complete the same task'

Example of a usability specification

- In the context of a library information system e.g.

Efficiency of first time use measured by retrieval time of specified information

'90 % of student users without any prior training should be able to retrieve the shelf number of a book, given the author and the title within 10 times the length of time an expert user would take to complete the same task'

Statement of usability attribute

Example of a usability specification

- In the context of a library information system e.g.

Efficiency of first time use measured by retrieval time of specified information

'90 % of student users without any prior training should be able to retrieve the shelf number of a book, given the author and the title within 10 times the length of time an expert user would take to complete the same task'

Statement of how it will be measured

Example of a usability specification

- In the context of a library information system e.g.

Efficiency of first time use measured by retrieval time of specified information

'90 % of student users without any prior training should be able to retrieve the shelf number of a book, given the author and the title within 10 times the length of time an expert user would take to complete the same task'

statement of the subset of users to which the specification applies

Example of a usability specification

- In the context of a library information system e.g.

Efficiency of first time use measured by retrieval time of specified information

'90 % of student users without any prior training should be able to retrieve the shelf number of a book, given the author and the title within 10 times the length of time an expert user would take to complete the same task'

statement of criteria that will represent attainment of the specification

Example of a usability specification

- In the context of a library information system e.g.

Efficiency of first time use measured by retrieval time of specified information

'90 % of student users without any prior training should be able to retrieve the shelf number of a book, given the author and the title within 10 times the length of time an expert user would take to complete the same task'

statement of pre-conditions of measurement

Impact Analysis (Gilb)

- Total time = time spent on task + time spent in error
- Identify errors made by subjects completing trial
- Estimate amount of time spent on each error
- Time saving provides estimate of *benefit* in designing a given error
- Benefit can now be offset against the *cost* of removing it (in development time)

A Sample metric – Work Speed

- Uniform user performance estimate measure for speed of using a bench task

$$S = \frac{P * C}{T}$$

- T= time user spent on task
- P= percentage of task completed
- C = a time unit constant
- S = user's performance score

Version 1: Initial measure – base points

	Time taken	% of task completed	Work speed	Evaluation rating
P1	21:82	100	14.3	2.8
P2	54:34	87	4.8	1.2
P3	25:28	94	11.1	1.6
P4	54:38	47	2.6	1.6
P5	48:26	94	7.8	2.0
P6	14:46	100	29.5	1.8
Mean	35:96	87	18.8	1.7

Analysis of Version 1 video tapes

problem	P1	P2	P3	P4	P5	P6
Window positioning	0:38		0:21	6:51	0:19	1:42
Menu choke <i>off b_y l</i>	0:17		0:14			
Confused between menus	0:19		0:37		0:58	
Print origin off			0:39	0:23		0:22
Move window before login		0:46			0:11	
Attaching keyboard		6:41		4:44		
Help window obscured		0:39		3:02		
Pressing near border					7:39	0:40
Inside/outside window			0:23			
Click/Press confusion			1:38			
CTRL-S not light LED				3:46		
Deleting windows		6:53				
Get menu when move window	0:55					
<i>Total problem time</i>	<i>2:09</i>	<i>14:50</i>	<i>3:44</i>	<i>18:46</i>	<i>8:58</i>	<i>2:44</i>
<i>Total task time</i>	<i>21:02</i>	<i>54:34</i>	<i>25:20</i>	<i>54:38</i>	<i>40:26</i>	<i>14:40</i>

Estimates of work speed improvements

	Measured V1 work speed	Estimated increment	New (estimated) work speed
P1	14.3	1.6	15.9
P2	4.8	1.7	6.5
P3	11.1	1.9	13.8
P4	2.6	1.3	3.9
P5	7.8	1.9	8.9
P6	29.5	4.6	25.1
<i>Mean</i>	<i>19.9</i>	<i>2.2</i>	<i>12.2</i>

Expect about 20% improvement

Problem ranked by impact

<i>Problem</i>	<i>Relative Impact on Initial use</i>	<i>Software change made ?</i>
Window Positioning	32%	Yes
Pressing near border	23%	Yes
Attaching keyboard	7%	Yes
Print origin off	7%	No
Deleting windows	6%	Yes
Click/press confusion	6%	No
Confused between menus	5%	Yes
Get menu when move window	5%	No
Menu choice off by 1	3%	Yes
CRTL-S not light LED	2%	No
Inside/outside window	2%	No
Help window obscured	1%	Yes
Move window before login	0%	Yes

Version 2 Initial Usability

	Time taken	% of task completed	Work speed	Evaluation rating
P7	18:86	100	16.6	8.8
P8	25:22	93	11.8	1.9
P9	15:48	94	17.8	1.4
P10	21:00	93	13.3	1.8
P11	46:28	100	7.4	2.9
P12	18:26	100	16.3	1.8
V2		97	13.7	1.3
Mean	23:18			
V1		87	18.8	1.7
Mean	35:86			

30% improvement in mean work speed compared with version 1

Impact analysis and eye pointing

- Very useful approach to discover the sources of user difficulties and how much improvement can be expected from design changes
- We've used this to compare eye and head pointing, and to assess the likely impact of changes made to eye pointing

Strengths of Usability Engineering

- agreeing on a definition of usability
- setting this definition in terms of metrics and usability goals
- putting usability on a par with other usability goals
- providing a method for prioritising usability problems

and Weaknesses....

- assumption that usability that usability can be operationalised
- requirement that practitioner is familiar with laboratory methods
- cost of conducting usability tests
- unnaturalness of testing environment

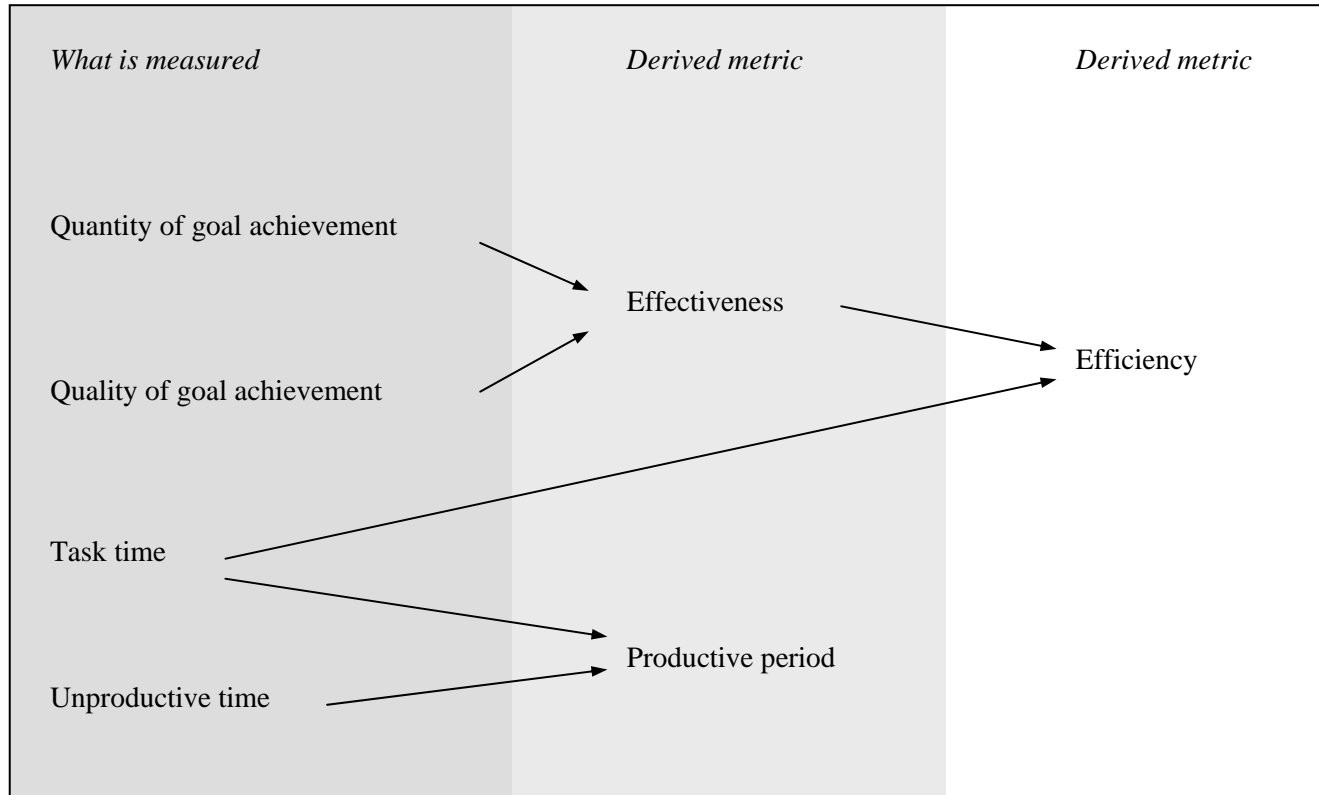
'Laboratory' vs. Real World

- work context
 - e.g. lab. test - 6 pages
 - normal work - reports of over 30 pages
- time context
 - lab test - time when task will be completed is prescribed
 - real world - individual has some control over scheduling
- motivational context
 - clear differences between lab and real world
- social context
 - lab test - no support
 - real world - social network of support - 'ask my friends'

MUSIC

- Metrics for Usability Standards in Computing
- Project undertaken in UK during 1990's (Bevan and MacLeod)
- **Quantity** measures – proportion of task goals represented in output
- **Quality** measures – degree to which output achieves task goals
- **Duration** measures –task, help, search, snag
- **Count** measures – number of activities, number of productive activities

Measures and metrics (MUSiC)



Sample metrics

- Task Effectiveness = $(\text{Quantity} * \text{Quality})/100$
- User Efficiency = $\text{Effectiveness}/\text{Task time}$
- Relative User Efficiency = $(\text{User Efficiency}/\text{Expert Efficiency}) * 100\%$
- Productive Period = $(\text{task time} - (\text{helptime} + \text{search time} + \text{snag time})/\text{task time}) * 100\%$
- Productive Activity Level (PAL) = $(\text{Number of wholly productive activities}/\text{number of activities undertaken}) * 100\%$

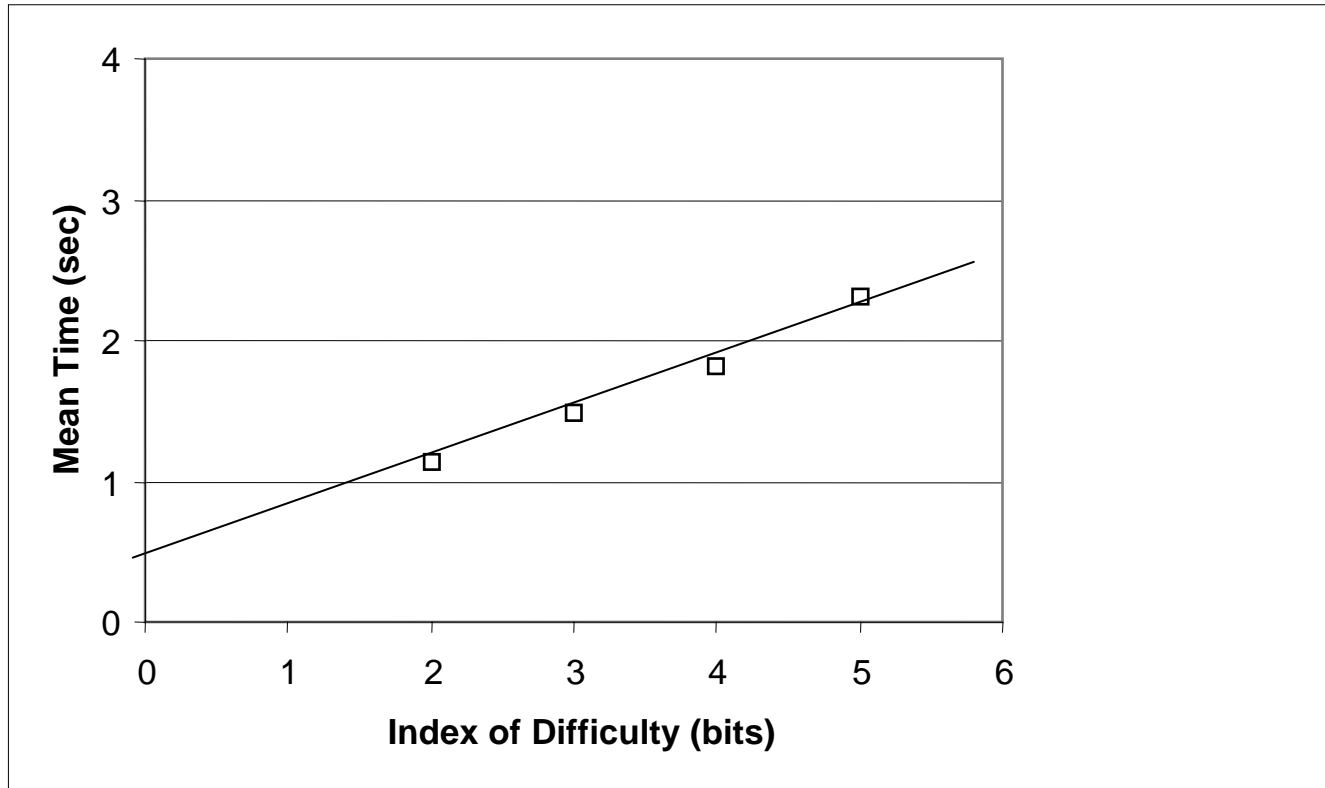
Comments on MUSiC metrics

- Very arbitrary (although not a bad thing in itself)
- Project never validated these metrics properly
- “Value of the metrics lies in the fact that the process of generating the metrics leads the discovery of usability issues” – or words to that effect....

Usability of Pointing Devices (ISO9241/9)

- ISO standard defines evaluation procedures for measuring performance, effort and comfort
- Variety of tasks (including tapping tasks)
- **Performance** : Throughput = ID_e/MT bits/sec
 - MT = mean movement time
(for all trials with same conditions)
 - $ID_e = \log_2(D/W_e + 1)$ Fitts Index of Difficulty
- **Effort** – biomechanical measurement of muscle load ☹️
- **Comfort** - device assessment questionnaire

Fitts Law



- $MT = a + b(ID)$ bits/sec
- Throughput = $1/b =$ Fitts Index of Performance

Throughput as a usability metric

- Many instances of using Fitts IP to characterise device performance - Mackenzie (1993)
- Poor consistency between independent estimates of IP for the 'same' device.
- Better consistency between the ratio of scores obtained from two similar devices within the same study

$$\text{mouse}_{(\text{study1})} / \text{joystick}_{(\text{study1})} = \text{mouse}_{(\text{study2})} / \text{joystick}_{(\text{study2})}$$

Why use gaze data in usability studies?

- Suppose a user spends longer than expected looking at a page in the course of achieving a goal
- Reasons for this are difficult to discover...
 - Observation of user gives coarse-grained information
 - Verbal protocols likely to interfere with search behaviour
 - User will not be able to recall detail of visual search accurately in post-trial walk-through
 - Logging of pointer position gives little information about what the user is looking at, or thinking

Benefits of Eye Gaze Analysis

- Can answer the following questions:
 - How do users look at a web page?
 - What attracts their attention first?
 - How long do they look at items?
 - Do they look at important features?
 - Do they miss links?
- Complements expert reviews and user trials
- Reveals the process by which web sites are comprehended and navigated
- Can show what users look at while the page is loading if the link is slow

Two similar sites

- Rock, Kitchen Harris are an advertising agency based in Leicester with a new division FourthFloor that undertakes web design commissions
- Both parts of the company have web pages – quite different in style
- FourthFloor pages can be reached from RKH's home page
- Purpose of both sites – to advertise and promote the professional services of both parts of the company
 - www.rkh.co.uk
 - www.fourthfloor.co.uk


Task scenario - 1

- You are responsible for commissioning a website for a company within the food industry. Find out whether RKH provide a web-site design and implementation service.
- If they have, look for and view at least one site they produced for a company within the food industry, if one exists. Record one significant item of information about the site you have found.

©DMU HCI Research Group - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Address <http://www.rkh.co.uk/> Go



rockkitchenharris

- advertising
- design
- new media
- public relations
- clients
- contact us
- recruitment
- fourthfloor

drop one small **stone** into a pool
and the **ripples** will find the farthest shore

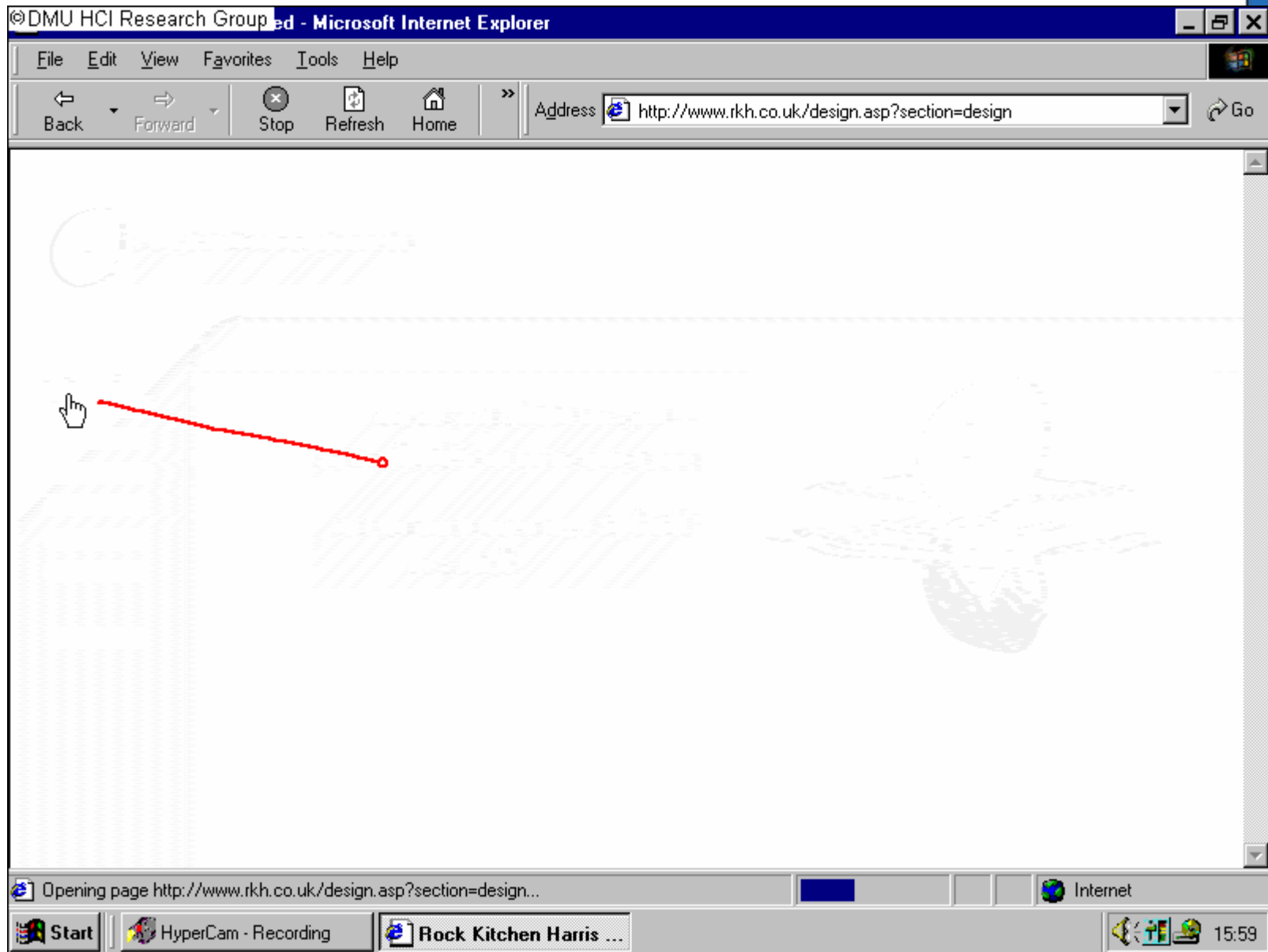
but if you want to make a bigger **splash**
use a **rock**

Rock Kitchen Harris Limited, based in Leicester, England was established as a full service Advertising Agency in 1982.

A member of the **Institute of Practitioners in Advertising**, we are proud of the experience we have developed in many areas including advertising, design, media, public relations, exhibitions, conferences and events. **fourthfloor@rkh** is expanding rapidly and producing websites and more using new technology to provide powerful marketing solutions.

[2 items remaining] Downloading picture <http://www.rkh.co.uk/images/bg.gif> Internet



Start HyperCam - Recording about:blank - Micros... 15:59



©DMU HCI Research Group - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Address http://www.rkh.co.uk/design_1.asp?section=design Go



advertising
design
new media
public relations
clients
contact us
recruitment
home
fourthfloor

Product design and identity for a ground breaking concept for the future of learning, if you want to know more, just read the papers.

(7 items remaining) Downloading picture <http://www.rkh.co.uk/images/bg.gif> Internet

Start HyperCam - Recording Rock Kitchen Harris ... 16:00

©DMU HCI Research Group ed - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Address <http://www.rkh.co.uk/design.asp?section=design> Go

rockkitchenharris design

advertising
design
new media
public relations
clients
contact us
recruitment
home
fourthfloor

design that runs deep

select a case study


http://www.rkh.co.uk/design_1.asp?section=design Internet

Start HyperCam - Recording Rock Kitchen Harris ... 16:00

©DMU HCI Research Group - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Address <http://www.rkh.co.uk/newmedia.asp?section=newmedia> Go



rockkitchenharris

- advertising
- design
- new media**
- public relations
- clients
- contact us
- recruitment
- home
- fourthfloor

select a case study

(8 items remaining) Downloading picture <http://www.rkh.co.uk/images/bg.gif> Internet

Start HyperCam - Recording Rock Kitchen Harris ... 16:00

Using gaze data

- Small .avi files can be produced from individual user trials to illustrate specific design points found from expert reviews and/or user trials
- Meaningful aggregation of fixation data from several subjects is difficult, easier with metrics produced from object analysis

Typical gaze patterns

- Work in progress to characterise
 - **reading** and scanning text (straightforward..?)
 - **confusion** (hunting rapidly over parts of the page and revisiting the same parts)
 - **distraction** (brief glances to a particular object)
 - **indecision** (gaze flicks between two candidate choices)
 - **complexity, ambiguity, incomprehension**

Conclusions

- Validity and reliability of any metric has to be established and understood – especially if the cost of obtaining the metric is high.
- Plenty of qualitative information on usability can be obtained from data for generating metrics
- Impact analysis offers the prospect of quantifying the benefits of making design changes.