User study : 2
In-game report

User data:

- Number of users: 14
- Male users: 11
- Female users: 3
- Interaction per user: 3
- Total interaction tasks: 84
  - NUI: 42 tasks
  - UNUI: 42 tasks

Distraction time:

Average distraction time :

- UNUI : 1.068 seconds
- NUI : 0.357 second

Average individual distraction time :
UNUI:

- Window: 1 second
- Mirror: 1.16 seconds
- Radio:
  - On/off: 1.2 seconds
  - Adjust volume: 1 second
  - Change track: 0.75 second

NUI:

- Window: 0.2 second
- Mirror: 0.58 second
- Radio
  - On/off: No distraction
  - Adjust volume: 1 second
  - Change track: No distraction

Conclusion based on distraction time analysis:

- UNUI is three times more distractive in comparison to NUI
- In UNUI mirror interaction is more distractive (due to two stage interaction)

- In NUI adjusting volume is more distractive (due to voice + interaction)

Collisions:

- **Total: 7**
  - During NUI: 2 (28% of total collisions)
  - During UNUI: 4 (57% of total collisions)
  - No interaction: 1 (15% of total collisions)

Probability of collision during interaction tasks:

- UNUI: 0.1 (1 in 10 interactions)
- NUI: 0.04 (1 in 21 interactions)

Other possibilities:

- During no interaction: 0.02 (1 in 42 interactions)
- Probability of no collisions during interaction: 0.84 (35 in 42 interactions)

Conclusion based on number of collisions:

- Interacting with UNUI has two times more probability of a collision than NUI
- Collision happens once in 10 interactions during interacting with UNUI
- Collision happens once in 21 interactions during interacting with NUI
- NUI interaction is much safer compared to UNUI interaction method.

False positives:

- During NUI: 3 times
- During UNUI: 7 times

Conclusion based on false positives:

- UNUI is twice prone to errors than NUI
- NUI is more reliable

Transformed Data

Ite												
1	2	3	4	5	6	7	8	9	10	11	12	13
2	2	3	2	-1	3	2	1	1	-2	2	1	-1
-1	0	2	1	-1	1	-1	-1	1	-2	0	2	1
2	3	3	3	-2	3	3	3	3	-3	3	3	3
2	1	2	2	-3	2	3	2	2	-2	1	2	-2
3	3	2	3	-2	2	3	3	-1	-3	3	3	3
0	2	1	2	-3	2	1	2	-2	-2	0	2	1
3	2	3	3	-3	3	2	2	3	-3	3	3	3
1	2	1	0	-2	1	3	2	1	-1	2	1	0
2	3	2	3	-2	3	3	3	2	-3	3	3	3
0	1	2	-2	0	1	-2	1	1	2	-2	0	-2
2	2	3	2	0	2	1	2	0	-1	2	3	-2
2	3	2	2	-2	3	2	2	3	-3	2	3	2
1	2	0	3	0	1	2	1	-1	0	0	2	2
-1	2	0	-2	0	0	-3	1	1	0	-3	0	-3

ms												
14	15	16	17	18	19	20	21	22	23	24	25	26
2	1	3	0	2	1	1	-1	1	2	2	2	0
1	2	1	-2	1	0	2	-2	1	0	2	1	-1
3	2	3	1	2	-2	2	-2	1	2	3	3	1
2	3	2	0	1	-1	2	-2	1	1	2	3	-2
3	2	3	1	3	-3	3	-3	3	3	3	3	-3
0	0	1	0	1	-1	1	-2	1	2	1	2	2
3	3	3	3	3	3	3	-3	3	3	3	3	-3
1	1	2	-2	1	-1	0	-1	1	2	1	1	1
2	3	2	2	3	-2	3	-3	2	3	3	3	2
-1	2	0	-2	-2	0	-2	-2	-2	1	-2	-2	0
2	3	3	2	3	3	3	-3	2	3	3	3	-3
1	2	1	1	3	1	2	-3	2	3	3	2	2
-1	-1	-1	1	2		2	-2	2	0	2	2	-1
1	0	-1	-3	-2	-3	0	-1	-2	3	0	0	3

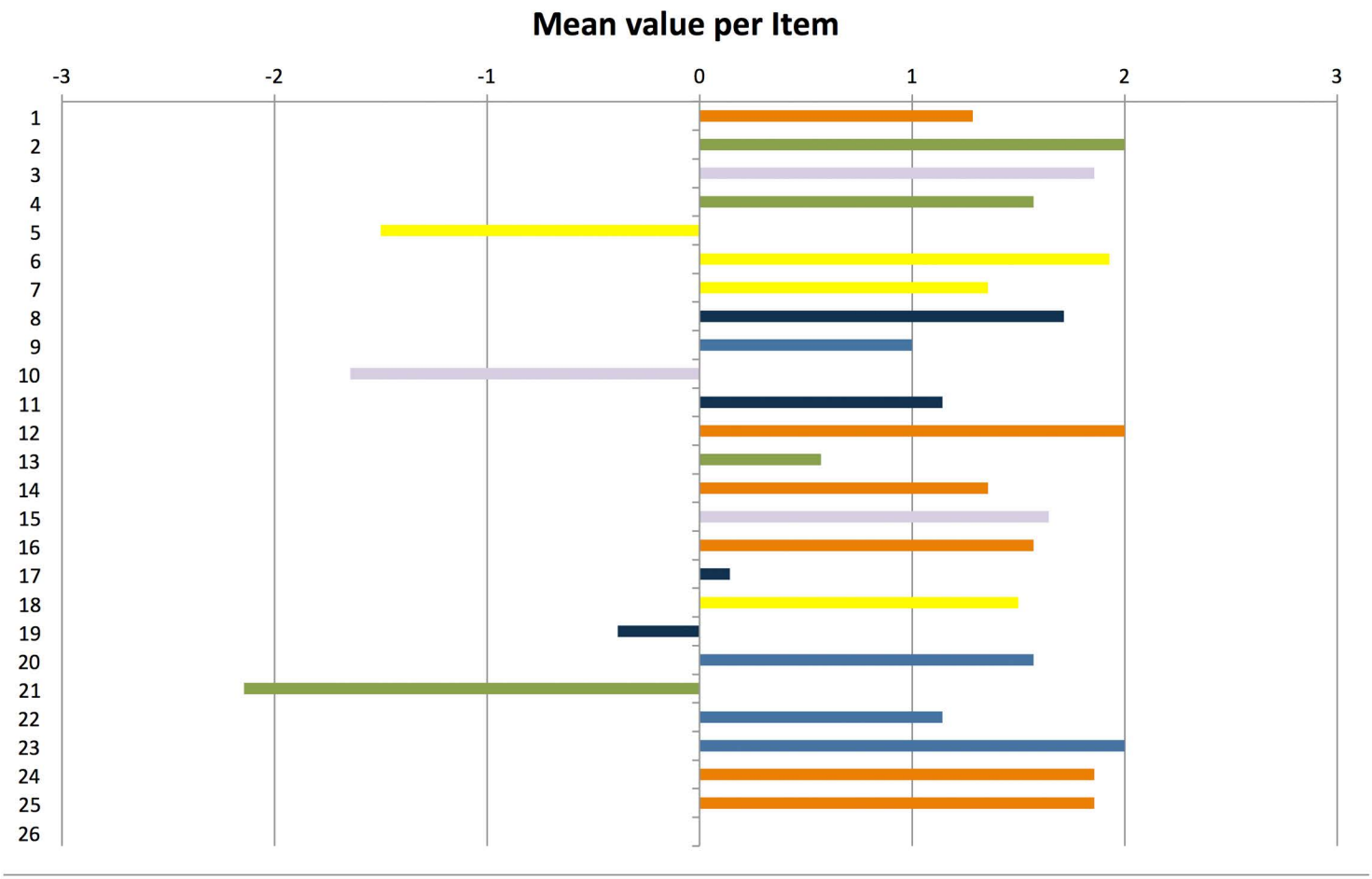
Skale means per person					
ATT	DUR	EFF	STEU	STI	ORI
2.00	0.50	1.25	1.00	1.50	0.50
1.00	0.00	1.00	-0.75	0.00	0.25
2.83	1.75	2.00	1.25	1.50	0.75
2.17	-0.25	1.50	0.50	0.75	0.25
3.00	1.50	2.00	1.00	1.50	-0.50
1.00	0.75	0.50	0.25	0.25	0.25
3.00	1.25	3.00	2.75	1.25	0.00
1.17	0.25	1.00	0.25	0.75	0.50
2.50	1.50	2.50	1.50	1.75	1.00
-0.83	-1.25	-0.50	-0.75	-0.75	1.50
2.67	-0.25	2.00	2.25	1.50	0.50
2.00	1.00	2.50	1.50	1.50	0.75
0.83	1.25	0.75	0.67	1.25	-0.50
-0.17	-1.00	0.50	-2.00	-1.25	0.75

Results

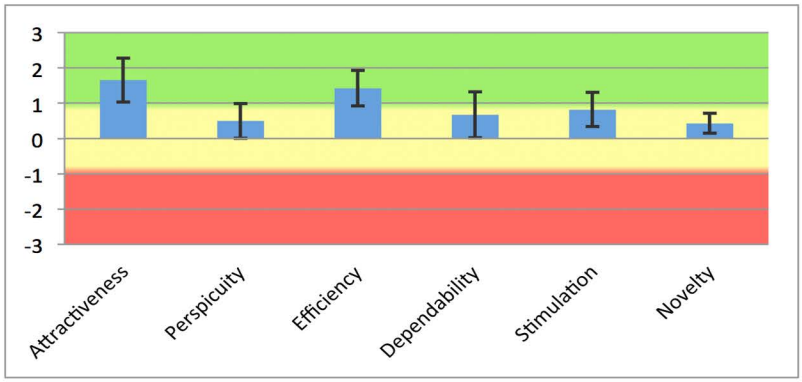
You can interpret the means of the scales. The UEQ does not produce an overall score for the user experience. Because of the construction of the quest the mean over all scales), since this value can not be interpreted properly. The values for the single items are listed to allow you to detect outliers in the e same scale this can be a hint that the item is misinterpreted (for example because of a special context in your evaluation) by a higher number of participa Values between -0.8and 0.8 represent a *neural evaluation* of the corresponding dimension, values > 0,8 represent a *positive evaluation* and values < -0, The range of the scales is between -3 (horribly bad) and +3 (extremely good). But in real applications in general only values in a restricted range will be o different opinions and answer tendencies, for example the avoidance of extreme answer categories, extremely unlikely to observe values above +2 or be Thus, even a quite good value of +1.5 for a scale looks from the purely visual standpoint on a scale range of -3 to +3 not as positive as it r the scale means. Use the figure with the reduced scale -2 to +2 if you communicate the results to persons that have not much knolwedge explain in detail how building mean values and answer tendencies influence the observed data.

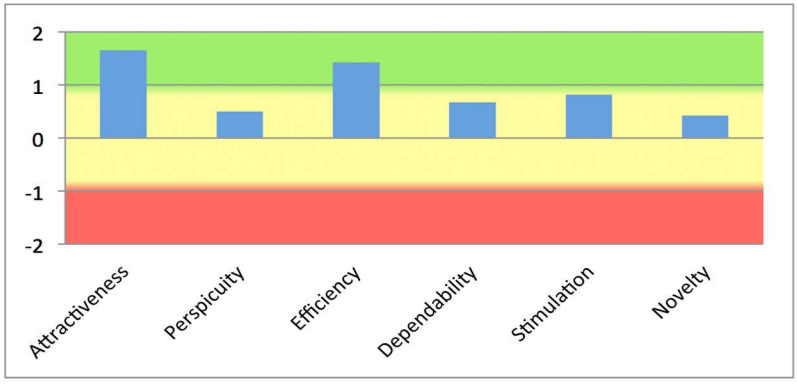
Text

Item	Mean	Variance	Std. Dev.	No.	Left	Right	Skale	
1	1.3	1.8	1.3	14	annoying	enjoyable	Attractiveness	
2	2.0	0.8	0.9	14	not understandable	understandable	Perspicuity	
3	1.9	1.1	1.0	14	creative	dull	Novelty	
4	1.6	3.0	1.7	14	easy to learn	difficult to learn	Perspicuity	
5	-1.5	1.3	1.2	14	ordinary	novel	Stimulation	
6	1.9	1.0	1.0	14	boring	exciting	Stimulation	
7	1.4	3.9	2.0	14	confusing	clearly structred	Stimulation	
8	1.7	1.1	1.1	14	unpredictable	predictable	Dependability	
9	1.0	2.5	1.6	14	fast	slow	Efficiency	
10	-1.6	2.2	1.5	14	rejecting	inventing	Novelty	
11	1.1	3.7	1.9	14	obstructive	supportive	Dependability	
12	2.0	1.2	1.1	14	good	bad	Attractiveness	
13	0.6	4.9	2.2	14	complicated	easy	Perspicuity	
14	1.4	1.8	1.3	14	unlikable	pleasing	Attractiveness	
15	1.6	1.6	1.3	14	usual	leading edge	Novelty	
16	1.6	2.1	1.5	14	unpleasant	pleasant	Attractiveness	
17	0.1	3.2	1.8	14	secure	not secure	Dependability	
18	1.5	2.9	1.7	14	motivating	demotivating	Stimulation	
19	-0.4	3.9	2.0	13	undemanding	challenging	Dependability	
20	1.6	2.1	1.5	14	inefficient	efficient	Efficiency	
21	-2.1	0.6	0.8	14	conservative	innovative	Perspicuity	
22	1.1	2.3	1.5	14	impractical	practical	Efficiency	
23	2.0	1.2	1.1	14	organized	cluttered	Efficiency	
24	1.9	2.1	1.5	14	attractive	unattractive	Attractiveness	
25	1.9	2.1	1.5	14	friendly	unfriendly	Attractiveness	



Scales	
Attractiveness	1.655
Perspicuity	0.500
Efficiency	1.429
Dependability	0.673
Stimulation	0.821
Novelty	0.429





Confidence intervals

Confidence interval (p=0.05) per item						
Item	Mean	Std. Dev.	N	Confidence	Confidence interval	
1	1.286	1.326	14	0.695	0.591	1.980
2	2.000	0.877	14	0.459	1.541	2.459
3	1.857	1.027	14	0.538	1.319	2.395
4	1.571	1.742	14	0.912	0.659	2.484
5	-1.500	1.160	14	0.608	-2.108	-0.892
6	1.929	0.997	14	0.522	1.406	2.451
7	1.357	1.985	14	1.040	0.317	2.397
8	1.714	1.069	14	0.560	1.154	2.274
9	1.000	1.569	14	0.822	0.178	1.822
10	-1.643	1.499	14	0.785	-2.428	-0.858
11	1.143	1.916	14	1.004	0.139	2.146
12	2.000	1.109	14	0.581	1.419	2.581
13	0.571	2.209	14	1.157	-0.586	1.728
14	1.357	1.336	14	0.700	0.657	2.057
15	1.643	1.277	14	0.669	0.974	2.312
16	1.571	1.453	14	0.761	0.811	2.332
17	0.143	1.791	14	0.938	-0.795	1.081
18	1.500	1.698	14	0.890	0.610	2.390
19	-0.385	1.981	13	1.077	-1.461	0.692
20	1.571	1.453	14	0.761	0.811	2.332
21	-2.143	0.770	14	0.404	-2.546	-1.739
22	1.143	1.512	14	0.792	0.351	1.935
23	2.000	1.109	14	0.581	1.419	2.581
24	1.857	1.460	14	0.765	1.092	2.622
25	1.857	1.460	14	0.765	1.092	2.622
26	0.000	0.000	0	#ZAHL!	#ZAHL!	#ZAHL!

for items and scales

Confidence intervals (p=0.05) per scale						
Scale	Mean	Std. Dev.	N	Confidence	Confidence interval	
Attractiveness	1.655	1.194	14	0.625	1.029	2.280
Perspicuity	0.500	0.951	14	0.498	0.002	0.998
Efficiency	1.429	0.963	14	0.504	0.924	1.933
Dependability	0.673	1.250	14	0.655	0.018	1.327
Stimulation	0.821	0.932	14	0.488	0.333	1.310
Novelty	0.429	0.541	14	0.283	0.145	0.712

Correlations of the items per scale and Cronbachs Alpha-Coefficient

Items that belong to the same scale should show in general a high correaltion. The Alpha-Coefficient (Cronbach, 1951) is a measure for the consistence of a scale. There is no generally accepted rule how big the value of the coefficient should be. Many authors assuem that a scale should show an alpha value > 0.7 to be considered as sufficiently consistent. However, from an methodological standpoint such a use of a cut-off criterium is not really well-founded (see for example Schmitt, N., 1996). Especially if you have only a small sample the value of the Alpha-Coefficient should be interpreted carefully.

If the value of the Alpha-Coefficient for a scale shows a massive deviation from a reasonable target value, for example 0.7, this can be a hint that some items of the scale are in the given context interpreted by several participants in an unexpected way. In such cases the corresponding scale should be interpreted very carefully.

Attractiveness	
Items	**Correlation**
1, 12	0.68
1, 14	0.68
1, 16	0.75
1, 24	0.70
1, 25	0.74
12, 14	0.57
12, 16	0.57
12, 24	0.90
12, 25	0.85
14, 16	0.84
14, 24	0.70
14, 25	0.70
16, 24	0.62
16, 25	0.66
24, 25	0.93
DK	0.73
Alpha	**0.94**

Perspicuity	
Items	**Correlation**
2, 4	0.45
2, 13	0.52
2, 21	-0.34
4, 13	0.73
4 , 21	-0.57
13, 21	-0.54
DK	0.04
Alpha	**0.15**

Efficiency	
Items	**Correlation**
9, 20	0.10
9, 22	0.00
9, 23	0.22
20, 22	0.87
20, 23	0.29
22, 23	0.23
DK	0.28
Alpha	**0.61**

Dependability	
Items	**Correlation**
8, 11	0.62
8, 17	0.59
8, 19	-0.24
11, 17	0.76
11, 19	0.22
17, 19	0.42
DK	0.39
Alpha	**0.72**

Stimulation	
Items	**Correlation**
5, 6	-0.57
5, 7	-0.62
5, 18	-0.41
6, 7	0.67
6, 18	0.75
7, 18	0.79
DK	0.10
Alpha	**0.31**

Novelty	
Items	**Correlation**
3, 10	-0.46
3, 15	0.78
3, 26	-0.44
10, 15	-0.41
10, 26	0.09
15, 26	-0.43
DK	-0.15
Alpha	**-1.04**

Benchmark: The measured scale means are set in relation to existing values from a benchmark data set products (business software, web pages, web shops, social networks).
The comparison of the results for the evaluated product with the data in the benchmark allows conclus

Please help to increase the data basis for the benchmark! If you use the UEQ to evaluate products it w product, the number of participants in your study and the scale means. We will of course handle these i

Scale	Mean	Comparisson to benchmark
Attractiveness	1.654761905	**Good**
Perspicuity	0.5	**Bad**
Efficiency	1.428571429	**Good**
Dependability	0.672619048	**Bad**
Stimulation	0.821428571	**Below Average**
Novelty	0.428571429	**Below Average**

:. This data set contains data from 4818 persons from 163 studies concerning different

ions about the relative quality of the evaluated product compared to other products.

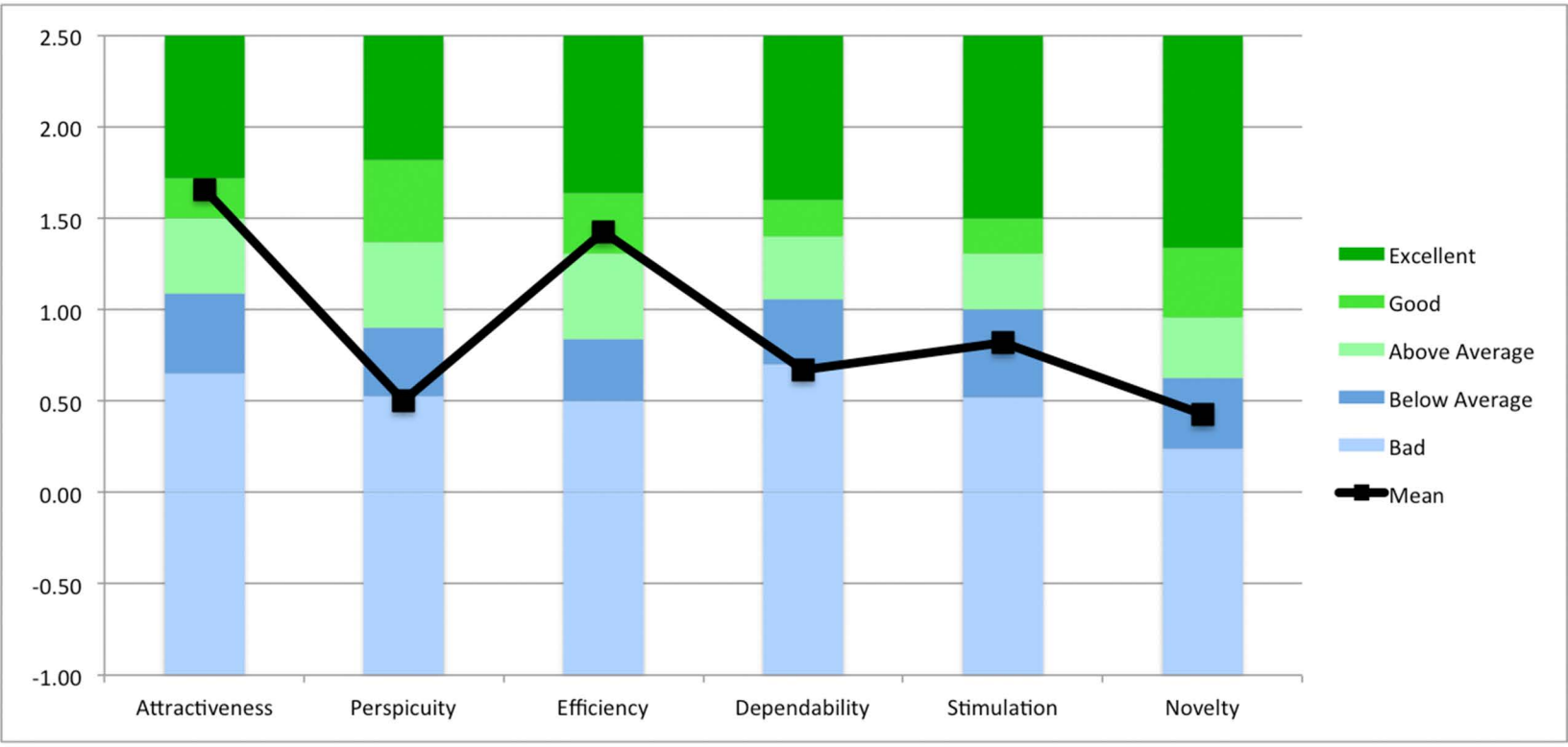
ould be quite helpful for us if you will share information concerning the type of
information absolutely confidential and will use it solely to improve the benchmark.

Interpretation

10% of results better, 75% of results worse
In the range of the 25% worst results
10% of results better, 75% of results worse
In the range of the 25% worst results
50% of results better, 25% of results worse
50% of results better, 25% of results worse

Graph to show the result relative to the benchmark

Scale	Lower Border	Bad	Below Average	Above Average	Good	Excellent	Mean
Attractiveness		-1.00	0.65	0.44	0.41	0.22	0.78	1.654761905
Perspicuity		-1.00	0.53	0.37	0.47	0.45	0.68	0.5
Efficiency		-1.00	0.5	0.34	0.47	0.33	0.86	1.428571429
Dependability		-1.00	0.7	0.36	0.34	0.2	0.9	0.672619048
Stimulation		-1.00	0.52	0.48	0.31	0.19	1	0.821428571
Novelty		-1.00	0.24	0.39	0.33	0.38	1.16	0.428571429



Interview Questions:

Personal interviews were conducted with the user to find his opinion on the system and also to know the difficulties that he faced personally while using the system. The following questions were asked in the Interview session.

1. Which method (Natural or Conventional) will you use if you have the choice of both in a car of your own ?
2. How was your overall experience about the system ?
3. What do you feel about the cognitive load of the proposed system ?
4. Did you face any difficulties in the system?
5. Any suggestions for Improvement ?

Conclusions based on the feedback from users:

- When we look into the ratio, 30 % of users chose natural,30% chose both whereas 40 % of users conventional.
- The Idea looks innovative, since it makes driving enjoyable. But it needs lot of improvements to adopt it into cars.
- Cognitive load is not a serious concern in general but some interactions are causing distractions at some particular instances. (e.g) when a user wants to adjust the side mirror to see something behind he has to adjust by seeing at the mirror as the user cannot see whether he could see the road properly without seeing the mirror itself, which causes lot of distraction.

Suggestions for Improvement:

- Position of sensors should be modified.
- Increase the session for each Interaction. It would be annoying to re-start the each interaction from start.
- Feedback icon should be closer to the eyesight to make driving more comfortable
- Since everyone is new, needs more training on the system to get acquainted to proposed methodology
- Leap should have more coverage area to operate the gestures freely.