

Lezione 10 17/04/2024

User tests

Quando si sviluppa un'applicazione, è necessario fare test. In informatica pensiamo più ai test funzionali sul software, ma questi non sono gli unici test da fare. Ci sono anche dei test che vanno fatti con gli utenti finali. Bisogna quindi fare **test di usabilità**. Dobbiamo andare a capire con l'utente se il nostro software è usabile, è piacevole.

Ogni soggetto si comporta in maniera differente, l'interazione con l'utente non può essere simulata, servono **test subjects** (devono essere simili a quelli che saranno gli utenti finali, rispetto al target). Quindi si può fare un ciclo di sviluppo che include i test con l'utente, prima di passare alla fase di sviluppo successiva.

Si prova a creare un software che sia fortemente iterativo, in real time, con grandi aspetti di immersione, quindi la parte di test con l'utente è estremamente importante su questi sistemi AR/VR.

Test steps

- **Preparazione:** prima che il partecipante inizi il test, si decide come fare l'esperimento, e poi si reclutano le persone. Questo include la fase dove si fornisce informazione riguardante al test al partecipante. Bisogna tenere in considerazione che se l'utente è a conoscenza del test, questo può avere un approccio diverso. Bisogna anche cercare di evitare di dire al test subject quello che stiamo cercando, limitandosi a dire quello che succederà a loro (per evitare che l'utente noti/non noti cose che avrebbe (o non avrebbe) notato in una situazione normale).
- **L'esecuzione del test.**
- Il **debrief**, dove possiamo fare domande al partecipante (o questionari).

La durata del test non dovrebbe superare i 45 minuti. Si può avere una fase di **pre test** con da 2 a 4 partecipanti per verificare la chiarezza delle istruzioni della fase di preparazione. Per fare il test è necessario firmare delle autorizzazioni di **consenso**.

Bias

Bisogna evitare di introdurre **bias** nell'esperimento, perché può falsare i risultati. Quindi dare troppa informazioni sul test, su quello che cerchiamo, se andiamo a fare una selezione di persone che appartengono ad uno stesso gruppo (le persone saranno difficilmente generalizzabili).

L'**effetto di Hawthorne** dice che il solo fatto che i soggetti sanno di far parte di un test cambia il loro comportamento e falsa i risultati. Quindi creare un ambiente calmo e rilassato può limitare questo effetto.

L'esecuzione del test deve essere controllata, i test devono essere fatti nello stesso modo (anche a livello di stress, condizioni ambientali, orario della giornata)

Variable types

L'idea è quella di individuare un'insieme di **variabili indipendenti**, andando a studiare l'effetto che il cambiamento di queste ha su un altro insieme di variabili. Per esempio se voglio studiare l'effetto di una grafica diversa nell'user interface, rispetto alla capacità dell'utente di interagire con il sistema. Quindi possiamo analizzare l'intuitività, il tempo che l'utente ci mette ad utilizzarla...

Quindi le **variabili indipendenti** sono quelle manipolate direttamente dal ricercatore, mentre le **variabili dipendenti** sono misurate.

Si parla di **studi quantitativi**, ovvero dove possiamo andare a fare delle statistiche su questi numeri, generalizzando la popolazione.

Se lo studio porta a dei risultati statistici significativi allora possiamo dire che un cambiamento nella variabile indipendente ha causato un cambiamento di una variabile dipendente.

Test design

Sono di due tipi:

- **within-group design**: ogni partecipante testa per una data variabile indipendente, diversi suoi valori. Per esempio nell'esempio precedente, potremmo far fare una task sull'interfaccia grafica A, e poi la stessa task all'interfaccia grafica B. Questo setup permette di avere meno soggetti perché possiamo fare più test, ma di contro l'utente tende a stancarsi di più (cyber sickness) ma soprattutto l'utente tende ad imparare quello che sta facendo, quindi quando utilizza l'interfaccia B avrà una conoscenza in più che può introdurre un bias.

- **between-group design**: ogni partecipante testa solamente una variazione della variabile indipendente. Sono necessarie più persone ma si evita il problema dell'influenza.

L'assegnamento delle persone ai vari gruppi va fatto in maniera randomica, per evitare di introdurre un bias nella selezione (come nel fare gruppi di soli uomini e soli donni, a meno che stiamo studiando proprio questa divisione).

Bisogna cercare di evitare il **learning effect**, per esempio facendo un task al di fuori dalla fase di valutazione, per far prendere confidenza all'utente con l'applicazione, e facendo partire l'esperimento dopo che l'utente ha capito come funziona il sistema.

Esempio

- Ex: We are testing two different interfaces (A and B), and we were interested in knowing how participants younger than 30 perform compared with older participants
- In this case we would **have two independent variables**:
 - Age, with 2 levels: under 30, over 30
 - Interface, with 2 levels: A and B
- For the study, we will recruit an equal number of participants in each age group
- Each participant (under or over 30) will perform an interaction using both A and B interfaces:
 - within-subjects w.r.t. the independent variable *Interface* (because each person sees both levels of this variable)
 - between-subjects w.r.t Age: one person can only be in a single age group (either under or over 30, not both)

Come collezionare i dati?

L'utente deve compilare un **questionario demografico**, per raccogliere alcune informazioni sull'utente (per esempio anche se è familiare con gli ambienti VR).

Può essere utile appuntarsi cosa dice l'utente a voce **durante il test**. Si possono anche collezionare dati qualitativi oltre a quelli descrittivi, per esempio quanti click fa, etc.

Si può cercare di dare dei **codici** ad alcune cose che ci aspettiamo che l'utente possa dirci (es cyber sickness) per poi valutare questi risultati statisticamente in modo quantitativo.

Per i **questionari** e le interview invece bisogna scegliere bene quali domande fare all'utente. Si possono usare questionari standard oppure ad hoc.

- ISONORM questionnaire for usability (Prümper 1993)
- AttrakDiff questionnaire for measuring user experience
- the QUIS questionnaire (Questionnaire for User Interaction Satisfaction)
- **VRUSE** (Kalawsky ,1999)
- **SSQ** (Simulator Sickness Questionnaire) Kennedy et al., (1993)
- NASA TLX
- questionnaires for assessing presence such as WS (Wittmer and Singer 1998) and SUS (Slater et al. 1994)

Scelte multiple e scale possono essere usate nei questionari per raccogliere dati quantitativi.

Quantitative data

- **Dati nominali:** sono dati che non possono essere ordinati (per esempio se l'utente è uomo o donna)
- **Dati ordinali:** sono dati che si possono ordinare, per esempio l'età. Non è detto sapere la differenza tra due valori, per esempio nel caso della Likert scale (quella che ha d'accordo, neutrale, non d'accordo...)
- **Dati razionali:** oltre a poterli ordinare, si può anche a capire la distanza matematica tra due valori, mettendoli quindi uno in relazione all'altro.

Per questo tipo di dati so può fare anche la parte di valutazione statistica. Nel fare queste analisi bisogna innanzitutto capire se ci sono delle distribuzione normale, per poter applicare test come la media aritmetica e la deviazione standard. Se però la popolazione non ha una distribuzione normale ci possono essere problemi.

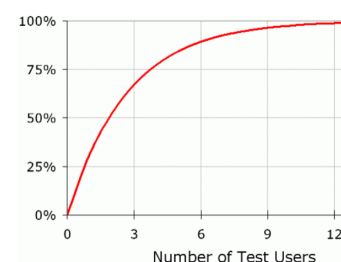
Numero di utenti

Quanti utenti servono per fare i test?

Se bisogna fare un **test iniziale di usability**, allora i risultati devono essere statisticamente significativi per ricavarne informazioni. Quindi servono almeno 20-40 partecipanti.

Altrimenti se stiamo cercando di identificare velocemente i **problemi principali** di usability possiamo seguire la **5 users rule**.

In generale bisogna avere almeno 3 partecipanti per catturare una diversità nel test group, o anche fino a



7 per assicurarsi di trovare quasi tutti i problemi di usabilità nel prodotto.

Ethical, Social and Legal Aspects of VR/AR

Quando si progetta un'applicazione di AR o VR, questa può avere parecchio impatto sull'utente, sia a livello di cyber sickness, ma anche sul modo di pensare e percepire se stesso. Su alcuni soggetti, queste applicazioni possono avere un impatto notevole.

Più l'utente può interagire attivamente, più stimoli fisici ci sono e più l'esperienza sarà reale. Di conseguenza, l'esperienza può anche potenzialmente avere un effetto più forte sulla psiche e, spesso inconsciamente, sui pensieri, sulle emozioni e sul comportamento.

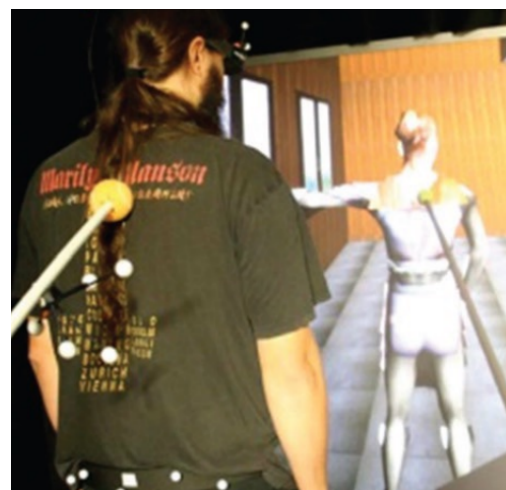
Alcuni tipi di applicazioni VR, vanno a sfruttare gli stereotipi, ma possono anche andare a cambiarli. Quindi l'utente può attribuire stereotipi, sia positivi che negativi, e gruppi di persone (di genere, di razza...). Se si mette l'utente un'avatar che appartiene ad un altro gruppo rispetto a quello reale della persona, la persona reagirà diversamente ad altri gruppi sociali, ma anche dopo l'esperienza tende ad avere un comportamento diverso (per esempio se metto un avatar femminile all'uomo, può succedere che il soggetto avrà una consapevolezza diversa rispetto a quella che aveva prima, l'utente può essere impattato).

Rubber hand e whole body illusions

Rubber hand illusion è il fenomeno dove l'utente si identifica non più con il suo corpo, ma con quello dell'avatar (almeno in modo parziale).

Questa trasformazione è creata per esempio quando un utente percepisce uno stimolo sulla sua schiena, ma lo può vedere soltanto sull'avatar.

Il **whole-body illusion** è quindi l'illusione di essere veramente nel mondo virtuale.



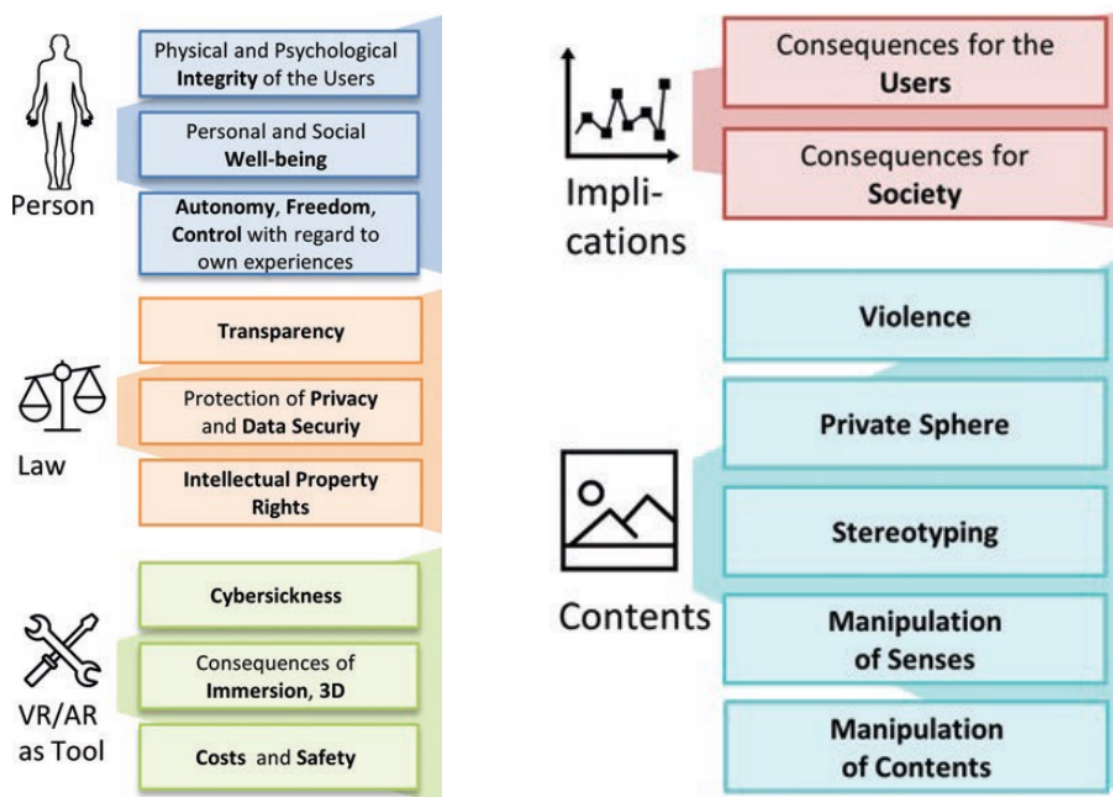
Escapismo

L'**escapismo** è il rischio che le persone fuggano dal reale per rifugiarsi nel virtuale. è un rischio quando il virtuale diventa estremamente realistico.

Questioni etiche

Bisogna sempre considerare i temi etici, che effetto avrà l'applicazione sulle persone.

Le leggi su queste applicazioni sono ancora molto generiche, non ci sono regolamentazioni sugli effetti dell'AR e VR sugli utenti.



THE CONTENT DESCRIPTORS

