

Μέθοδοι αξιολόγησης της διεπαφής

Δρ. Α. Κομνηνός
B.Sc. Hons, Ph.D.

31/5/2017 – Τμήμα Πληροφορικής - ΑΠΘ

Καλημέρα!

- Πτυχίο B.Sc. Honours Computer Studies (2001)
- Διδακτορικό Personal Predictive Internet Content Pre-caching for Mobile Devices (2005)
- Ερευνητικά ενδιαφέροντα (~70 peer-reviewed δημοσιεύσεις)
 - Mobile human-computer interaction
 - Context aware ubiquitous computing
- Τρέχουσα απασχόληση
 - Μέλος ΣΕΠ, ΕΑΠ
 - Course editor & author, Interaction Design Foundation (DK)
 - Συνεργαζόμενος μ/δ ερευνητής, ΤΜΗΥΠ Παν. Πατρών
- Προηγούμενη απασχόληση
 - Μέλος ΔΕΠ, Glasgow Caledonian University (2005-2013)
 - Μεταδιδάκτωρ ερευνητής, University of Strathclyde (2013-2016)
 - Έκτακτο μέλος ΕΠ Παν. Πατρών & ΑΤΕΙ Πατρών (2010-2015)

Η ανάγκη της αξιολόγησης

- Γενικότερα σκοπός της δημιουργίας ενός συστήματος για χρήση από ανθρώπους είναι η διευκόλυνση των δραστηριοτήτων που θέλουν να κάνουν
 - Επικοινωνία, διεκπεραίωση καθηκόντων, ψυχαγωγία, δημιουργία κλπ.
- Η χρήση από ανθρώπους προϋποθέτει κάποια έκταση αλληλεπίδρασης με το σύστημα, μέσω διεπαφών
- Η αλληλεπίδραση αυτή πρέπει να σχεδιάζεται με τέτοιο τρόπο ώστε να καθιστά το σύστημα
 - Χρήσιμο
 - (+) Εύχρηστο
 - (=) Χρησιμοποιούμενο!
- Για να γνωρίζουμε αν ο σχεδιασμός μας κατορθώνει τα ανωτέρω, χρειάζεται κάποιου είδους **αξιολόγηση**.

Τι αξιολογούμε;

- Παλαιότερα η έννοια της ευχρηστίας των διεπαφών ήταν στενά συνυφασμένη με την **απλότητα της χρήσης**.
 - Εστίαση στο γνωσιακό επίπεδο και σχέσεις αιτίασης μεταξύ διεπαφής και ανθρώπινης επίδοσης
 - Αν οι ανθρώπινες γνωσιακές ικανότητες είναι δεδομένες τότε μια διεπαφή έχει έμφυτη την έννοια της ευχρηστίας σε δυικό επίπεδο (είναι ή δεν είναι εύχρηστο).
 - Ψυχολογικού τύπου πειράματα χρησιμεύουν στην ανακάλυψη, έκφραση και επικύρωση κανόνων και προτύπων σχεδιασμού.

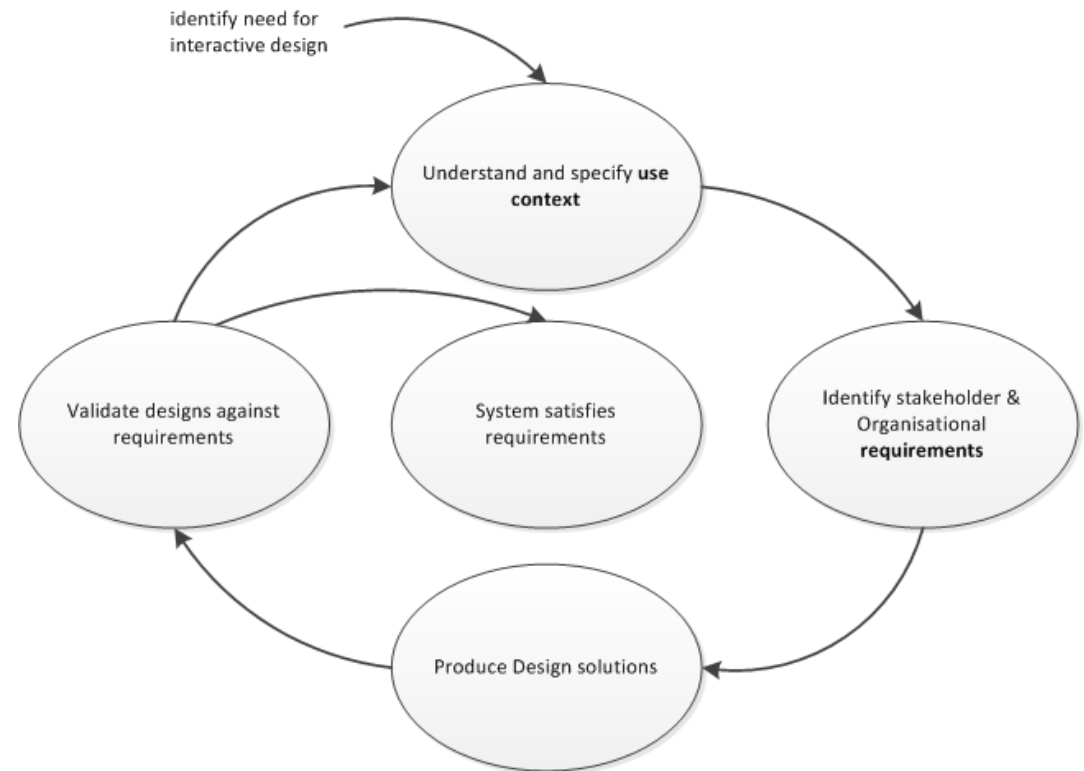
Τι αξιολογούμε;

- Αργότερα έγινε κατανοητό ότι η έννοια της ευχρηστίας δεν είναι δυική ιδιότητα του συστήματος αλλά εξαρτάται στενά από το πλαίσιο χρήσης (Rogers et al. 1994, Cockton 2004).
 - Οι γνωσιακές ικανότητες διαφέρουν **μεταξύ ατόμων**, και στο ίδιο άτομο **μεταξύ πλαισίων χρήσης**.
 - Η ευχρηστία είναι μια **αναδυόμενη ιδιότητα** της διεπαφής, που εξαρτάται
 - Από τα χαρακτηριστικά και την ποιότητα της διεπαφής
 - Από το ποιος τη χρησιμοποιεί
 - Τι προσπαθούν να κάνουν με αυτή
 - Πότε τη χρησιμοποιεί
 - Που τη χρησιμοποιεί
 - ...
- Εν τέλει, αξιολογούμε την **αξία** που προσφέρει ένα σύστημα σε συγκεκριμένα πλαίσια χρήσης

} Πλαίσιο (Context)

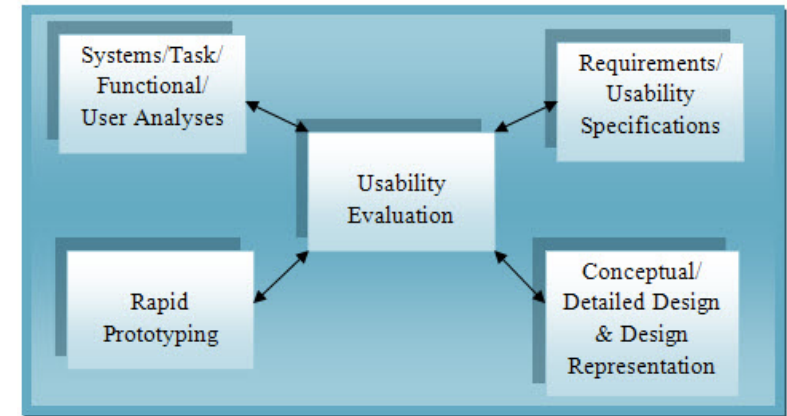
Πότε γίνεται η αξιολόγηση;

- Εδώ και δεκαετίες έχει εγκαταλειφθεί το παραδοσιακό μοντέλο ανάπτυξης συστημάτων
- Για την ανάπτυξη ανθρωποκεντρικών συστημάτων, χρησιμοποιούμε επαναληπτικά μοντέλα ανάπτυξης
- Ένα από τα πιο γνωστά πρότυπα είναι το ISO 9241-210 Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems



Πότε γίνεται η αξιολόγηση

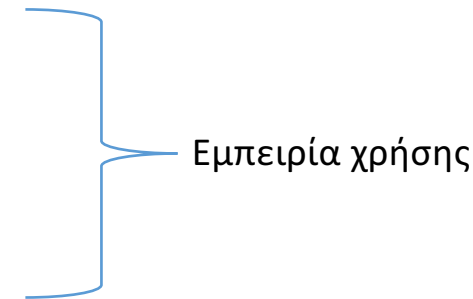
- Το πρότυπο βασίζεται στις εξής παραδοχές
 - Είναι σχεδόν αδύνατο για τους χρήστες να περιγράψουν επαρκώς τι χρειάζονται [επανάληψη μέχρι να καταλάβουμε]
 - Η έννοια της ευχρηστίας καλύπτει όλη την εμπειρία χρήσης και όχι μόνο την απλότητα [αντιληψιμότητα, συναισθήματα κλπ.] (βλ, και Cockton 2004)
 - Η ομάδα σχεδιασμού πρέπει να είναι διεπιστημονική [πολλαπλές οπτικές γωνίες]
- Η αξιολόγηση γίνεται σε όλα τα στάδια ανάπτυξης
 - Για να κατανοήσουμε τις ανάγκες και επιθυμίες των χρηστών / της επιχείρησης κλπ
 - Για να παράγουμε ιδέες σχεδιασμού
 - Για να υλοποιήσουμε τις ιδέες
 - Για να μελετήσουμε τις (μακροχρόνιες) επιπτώσεις χρήσης



Hix & Hartson 1993

Είδη αξιολόγησης

- Αξιολογούμε την αξία
 - Ευκολία κτήσης – εκμάθησης
 - Ευκολία χρήσης
 - Συναισθήματα
 - Ενσωμάτωση στην πρακτική / πραγματική ζωή
- Δεν υπάρχει μια μεθοδολογία ή τεχνική που να καλύπτει όλα τα στάδια
- Κάθε στάδιο χρειάζεται άλλου είδους τεχνικές, ή τις ίδιες αλλά προσαρμοσμένες και εφαρμοζόμενες διαφορετικά
- Κάθε οπτική γωνία (επιστήμη) προσφέρει τις δικές της τεχνικές
- Τα αποτελέσματα από τη σύνθεση των αποτελεσμάτων της εφαρμογής πολλαπλών τεχνικών σε κάθε στάδιο οδηγούν σε μια πληρέστερη εικόνα



Ταξινόμια ΜΕΑΞ Ι – Τύπος μεθόδου

• Τύπος μεθόδου

- Αξιολόγηση από ειδικούς
- Έρευνα πληθυσμού
- Αξιολόγηση στο εργαστήριο
- Αξιολόγηση στο πεδίο

• Είδος έρευνας

- Ποσοτική ανάλυση
- Ποιοτική ανάλυση

• Είδος χρήστη / αξιολογητή

- Απλοί χρήστες
- Ειδικοί

Ταξινόμια ΜΕΑΞ ΙΙ – Ανάπτυξη & εμπειρία χρήσης

- Στάδιο ανάπτυξης
 - Σύλληψη ιδέας
 - Ανάλυση
 - Σχεδιασμός
 - Ανάπτυξη πρωτοτύπων
 - Τελικό προϊόν

- Χρονικοί περιορισμοί
 - 1 ημέρα
 - 1 εβδομάδα
 - Μήνες

- Περίοδος εμπειρίας
 - Στιγμαία εμπειρία εφαρμογής
 - Επεισόδιο εμπειρίας εφαρμογής
 - Συνολική εμπειρία εφαρμογής

Συνηθέστεροι παράγοντες

- Με χρήστες
 - Από τη δεκαετία του 1980, ως μια παραλλαγή πειραμάτων γνωσιακής ψυχολογίας με εξαρτημένες μεταβλητές (το σύστημα είναι η ανεξάρτητη σταθερά) (Dumas & Redish 1993)
- Χωρίς χρήστες
 - Από τη δεκαετία του 1980 με χρήση μοντέλων, π.χ. GOMS (John & Kieras 1996)
 - Από τη δεκαετία του 1990, με την καθιέρωση των «εκπρωτικών μεθόδων» χωρίς χρήστες ή περιορισμένη συμβολή χρηστών (Nielsen 1994)

Dumas, Joseph S. and Redish, Janice C. (1993): A Practical Guide to Usability Testing. Norwood, NJ, Intellect

John, Bonnie E. and Kieras, David E. (1996): Using GOMS for User Interface Design and Evaluation: Which Technique?. In ACM Transactions on Computer-Human Interaction, 3 (4) pp. 287-319

Nielsen, Jakob (1994): Enhancing the explanatory power of usability heuristics. In: Plaisant, Catherine (ed.) Conference on Human Factors in Computing Systems, CHI 1994, Boston, Massachusetts, USA, April 24-28, 1994, Conference Companion 1994. p. 210

Συνηθέστεροι παράγοντες

- Χωρίς χρήστες
 - Πριν την διεξαγωγή πειραμάτων εργαστηρίου
 - Εμπλοκή ειδικών αξιολογητών
 - Χρήση ευρεστικών μεθόδων και κριτηρίων
 - Χρήση αναλυτικών μοντέλων (π.χ. KSPC)
- Με χρήστες
 - Σε ελεγχόμενο περιβάλλον (εργαστήριο)
 - Φθηνότερο, ελεγχόμενο, κατάλληλο για συγκεκριμένα στοιχεία εφαρμογής, βραχυχρόνιο
 - Στο πεδίο (field, in-the-wild)
 - Δυσκολότερη προσέλκυση, ανεξέλεγκτο, πιο ρεαλιστικό, μακροχρόνιο
- Είδη δεδομένων
 - Ποσοτικά δεδομένα (χρόνος εκπλήρωσης, ακρίβεια, σφάλματα, αριθμός πλήκτρων)
 - Ποιοτικά δεδομένα (υποκειμενικές απόψεις, σχόλια, κριτικές)

Χωρίς χρήστες

- Συστηματική επιθεώρηση για την ανακάλυψη σφαλμάτων / προβλημάτων και επίπεδου συμμόρφωσης με σχεδιαστικούς κανόνες
 - 3-5 επιθεωρητές (ειδικοί ή και χρήστες)
 - 1 ειδικός ~41% των προβλημάτων, 1 διπλά ειδικός ~60%, 1 απλός χρήστης ~22%, 5 ειδικοί ~75%
 - Ο καθένας μόνος του
 - Ανίχνευση και κατηγοριοποίηση σημαντικότητας προβλήματος
 - Σύγκριση σημειώσεων

Προβλήματα χωρίς χρήστες

- Ποιο σύνολο κανόνων / κριτηρίων θα χρησιμοποιηθεί;
 - Πόσο υποκειμενικές είναι οι αξιολογήσεις;
 - Πόσο ειδικοί είναι οι ειδικοί;
 - Πως επιλύονται διαφορές;
-
- Δεν πρόκειται για επιστημονική μέθοδο, αλλά για μια χρήσιμη πρακτική με εφαρμογή σε πρωταρχικά στάδια.

Ελεγχόμενα πειράματα

- Πιο επιστημονική μέθοδος!
- Σκοπός η διερεύνηση υποθέσεων σχετικά με την
 - Επίδοση των χρηστών
 - Ικανοποίηση των χρηστών
 - Συναισθήματα ή άλλες ψυχολογικά κατασκευάσματα
- Απαιτούνται
 - Ο σωστός σχεδιασμός του πειράματος για την απομόνωση μεταβλητών
 - Πλήθος συμμετεχόντων, ισοστάθμιση, τυχαιότητα, καθορισμός μετρικών και μεταβλητών, περιβάλλον
 - Οικολογική εγκυρότητα (ρεαλισμός)
 - Συμμετέχοντες, περιβάλλον, σενάριο, δραστηριότητες
 - Στατιστική ανάλυση / επεξεργασία δεδομένων

Πείραμα στο εργαστήριο ή στο πεδίο;

- Kjeldskov, J., Skov, M. B., Als, B. S., & Høegh, R. T. (2004, September). Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In *International Conference on Mobile Human-Computer Interaction* (pp. 61-73). Springer Berlin Heidelberg.
- Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., & Kankainen, A. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability studies*, 1(1), 4-16.
- Nielsen, C. M., Overgaard, M., Pedersen, M. B., Stage, J., & Stenild, S. (2006, October). It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles* (pp. 272-280). ACM.

Δε χρειάζεται το πεδίο I

- Η αξιολόγηση στο εργαστήριο ανακάλυψε το ίδιο πλήθος σφαλμάτων με το πεδίο
- Η έρευνα στο πεδίο εισήγαγε μεγάλη απώλεια ελέγχου
- Η έρευνα στο εργαστήριο ήταν εφικτό να αναπαράγει προβλήματα που σχετίζονται με το πλαίσιο χρήσης το ίδιο καλά με το πεδίο

Δε χρειάζεται το πεδίο II

- Τα προβλήματα «ρεαλισμού» δεν αποτέλεσαν εμπόδιο
- Υπάρχει πιθανότητα τα προβλήματα που ανακαλύφθηκαν στο εργαστήριο να μην έχουν πρακτική επίπτωση στην πραγματικότητα
- Προτείνεται τα πειράματα πεδίου να αφορούν άλλες φάσης του κύκλου ανάπτυξης (μακροχρόνια χρήση)

Χρειάζεται το πεδίο

- Τα συμπεράσματα των μελετών μπορεί να οφείλονται
 - Στο χαμηλό πλήθος συμμετεχόντων
 - Δε χρησιμοποιήθηκαν οι ίδιες τεχνικές συλλογής δεδομένων και στις 2 περιπτώσεις
 - Κάποιες από τις λεπτομέρειες της μεθοδολογίας ήταν αντικρουόμενες (λόγω συνθηκών)
- Διαφορετική μελέτη σε μια κινητή συσκευή και στις δύο συνθήκες έδειξε ότι
 - Τα προβλήματα ευχρηστίας που αφορούσαν γνωσιακές ικανότητες ή στυλ αλληλεπίδρασης έγιναν ορατά μόνο στο πεδίο
 - Η φύση της εργαστηριακής μελέτης αύξησε τεχνητά τα επίπεδα ενόχλησης και νοητικής απαίτησης από τους χρήστες
 - Το πεδίο ανέδειξε περισσότερα προβλήματα

Συμπέρασμα

- Είναι δύσκολο να γίνει αντικειμενική σύγκριση μεθόδων (διατηρώντας δηλαδή τη **μέθοδο** σαν *εξαρτημένη μεταβλητή*)
 - Οι μέθοδοι αξιολόγησης δεν είναι ορισμένες σε επίπεδο λεπτομέρειας που να επιτρέπει τη συνεπή εφαρμογή τους (Grey & Salzman 1998)
 - Υπάρχουν σημαντικές επιρροές από τους ίδιους τους ερευνητές (Hertzum & Jacobsen 2001)
- Χρειάζονται και τα δύο είδη μελέτης γιατί το καθένα μπορεί να έχει διαφορετική συνεισφορά στη διαδικασία αξιολόγησης
 - Με κάθε επιφύλαξη πάντως για τα αποτελέσματα!
- Ας δείξουμε ένα παράδειγμα

Ειδοποιήσεις!

- Δεκάδες ειδοποιήσεις καθημερινά στο κινητό μας
- Μελέτες (π.χ. Poppinga et al. 2014) που αφορούν την έκδοση ειδοποιήσεων υπό πλαίσιο ασχολούνται με την
 - Αυτόματη αναγνώριση της σημαντικότητας μιας ειδοποίησης
 - Αυτόματη αναγνώριση των καταλληλότερων στιγμών για την έκδοσή της
- Τι λείπει;
 - Η αυτόματη αναγνώριση του καταλληλότερου **τρόπου** έκδοσης μιας ειδοποίησης.
 - Για να βρούμε αυτό, πρέπει να κατανοήσουμε το βαθμό αντιληψιμότητας του κάθε τρόπου ειδοποίησης ή και συνδυασμού.

Στόχος

- Ποιος τρόπος παροχής ειδοποιήσεων είναι ο περισσότερο «αντιληπτός»;
 - Λιγότερο ενοχλητικές ειδοποιήσεις για μη-επείγοντα συμβάντα
 - Ειδοποιήσεις με ευαισθησία στο πλαίσιο του χρήστη
 - Αποφυγή διάσπασης προσοχής
 - Διατήρηση δυνατότητας μεταγενέστερης εμπλοκής με το συμβάν
- Ας μετρήσουμε το χρόνο μεταξύ έκδοσης μιας ειδοποίησης και της απόρριψής της (ή αποδοχής της).

Μελέτες in-the-wild

- Τέσσερις σχετικές μελέτες όπου καταγράφηκαν οι χρόνοι απόκρισης κατά την χρήση σε πραγματικές συνθήκες
 - Galud & Tesoriero (2015): Οποιοσδήποτε συνδυασμός αρκεί να έχει και ήχο
 - Mashhadi et al. (2014): Καμμία στατιστικά σημαντική διαφορά
 - Mehrotra et al. (2016): Χρόνος απόκρισης χαμηλότερος κατά σειρά δόνηση μόνο, ήχος μόνο, ήχος & δόνηση
 - Pielot et al. (2014): Χρόνος απόκρισης χαμηλότερος κατά σειρά με δόνηση μόνο, σιωπηλό, κανονικό (ήχος/δόνηση)
- Γιατί διαφέρουν τόσο τα αποτελέσματα;
- Μπορούμε να τα χρησιμοποιήσουμε ως βάση για το σχεδιασμό μας;

Gallud, J. A., & Tesoriero, R. (2015, August). Smartphone Notifications: A Study on the Sound to Soundless Tendency. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (pp. 819-824). ACM.

Mashhadi, A., Mathur, A., & Kawsar, F. (2014, September). The myth of subtle notifications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 111-114). ACM.

Mehrotra, A., Pejovic, V., Vermeulen, J., Hendley, R., & Musolesi, M. (2016, May). My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1021-1032). ACM.

Pielot, M., Church, K., & De Oliveira, R. (2014, September). An in-situ study of mobile phone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (pp. 233-242). ACM.

Ποιοτικά δεδομένα

- Ποιοτικά στοιχεία από άλλες μελέτες (Mehrotra et al. 2015, Okoshi et al. 2015, Poppinga et al. 2014, Shirazi et al. 2014) δείχνουν ότι ο χρόνος απόκρισης επηρεάζεται από το πλαίσιο του χρήστη
 - Ποιά εφαρμογή την εξέδωσε
 - Σχετική σημασία με βάση την τρέχουσα δραστηριότητα
 - Κοινωνική σχέση μεταξύ χρήστη και ατόμου από το οποίο προέρχεται
 - Ημέρα και ώρα
 - Συσκευή που εξέδωσε την ειδοποίηση
 - Προσωπικότητα του χρήστη

Mehrotra, A., Musolesi, M., Hendley, R., & Pejovic, V. (2015, September). Designing content-driven intelligent notification mechanisms for mobile applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 813-824). ACM.

Okoshi, T., Ramos, J., Nozaki, H., Nakazawa, J., Dey, A. K., & Tokuda, H. (2015, March). Attelia: Reducing user's cognitive load due to interruptive notifications on smart phones. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on* (pp. 96-104). IEEE.

Poppinga, B., Heuten, W., & Boll, S. (2014). Sensor-based identification of opportune moments for triggering notifications. *IEEE Pervasive Computing*, 13(1), 22-29.

Sahami Shirazi, A., Henze, N., Dingler, T., Pielot, M., Weber, D., & Schmidt, A. (2014, April). Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3055-3064). ACM.

Άλλες μελέτες

- Μελέτες από τον τομέα της προσοχής του χρήστη (user attention) και της γνωσιακής ψυχολογίας προτείνουν
 - Ο χρήστης (και όχι το σύστημα) πρέπει να λάβει την απόφαση τι να κάνει μόλις αντιληφθεί μια ειδοποίηση (Horvitz et al. 2003)
 - Επιπλέον παράγοντας είναι το εκτιμώμενο κόστος επαναφοράς στην τρέχουσα δραστηριότητα (Iqbal & Horvitz 2007)
 - Ο άνθρωπος ακολουθεί ένα σειριακό μοντέλο *sensation* → *perception* → *action* όπου η δράση αποφασίζεται μέσα από ένα πλαίσιο διατυπωμένων κανόνων (Ernst & Bühlhoff 2004)
 - Εναλλακτικό μοντέλο *perception* → *memory* → *judgement*, όπου η μνήμη μοντελοποιεί το συμβάν με βάση πρότερη γνώση (κανόνες) (Fu et al. 2014)

Horvitz, E., Kadie, C., Paek, T., & Hovel, D. (2003). Models of attention in computing and communication: from principles to applications. *Communications of the ACM*, 46(3), 52-59.

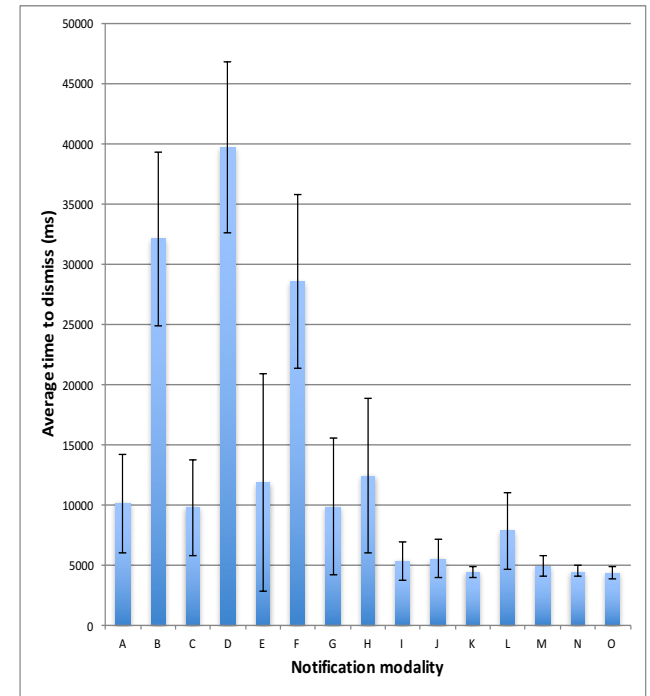
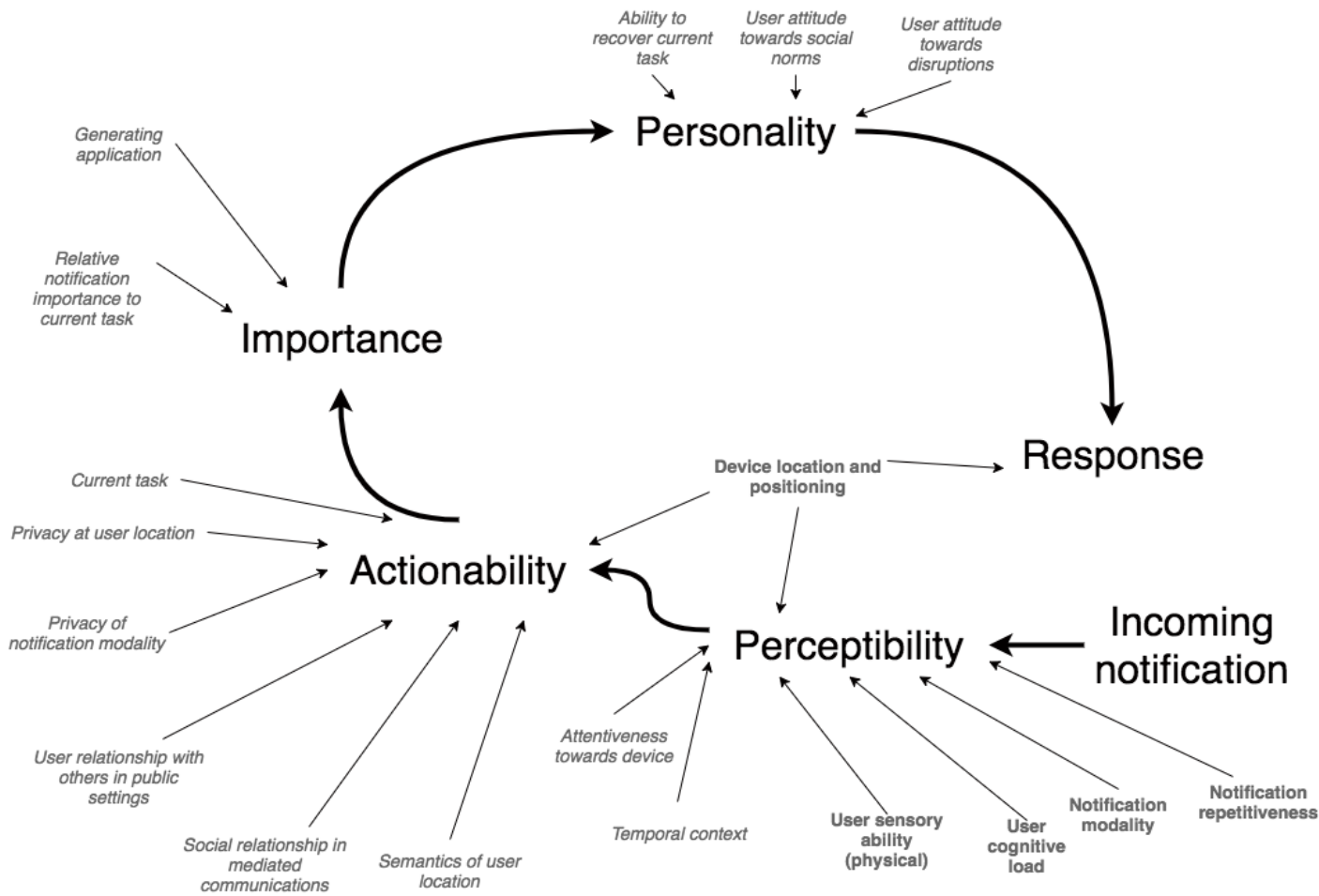
Iqbal, S. T., & Horvitz, E. (2007, April). Disruption and recovery of computing tasks: field study, analysis, and directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 677-686). ACM.

Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4), 162-169.

Fu, X., Cai, L., Liu, Y., Jia, J., Chen, W., Yi, Z., ... & Wu, C. (2014). A computational cognition model of perception, memory, and judgment. *Science China Information Sciences*, 57(3), 1-15.

Το σύνολο

- Τι λείπει;
 - Η αυτόματη αναγνώριση του καταλληλότερου τρόπου έκδοσης μιας ειδοποίησης.
 - Για να βρούμε αυτό, πρέπει να κατανοήσουμε το βαθμό αντιληψιμότητας του κάθε τρόπου ειδοποίησης ή και συνδυασμού.
 - Οι υπάρχουσες μελέτες βασίζονται μόνο σε ευρήματα από πειράματα στο πεδίο, ή σε ποιοτικά δεδομένα
 - Δεν είναι πλήρως κατανοητός ο μηχανισμός αντίληψης και λήψης απόφασης
 - Για να κατανοήσουμε το μηχανισμό, πρέπει να απομονώσουμε το χρήστη από το περιβάλλον του και να παρέμβουμε **πειραματικά με ελεγχόμενο τρόπο.**



Η συνέχεια σύντομα..

- Ευχαριστώ!
- Απορίες;
- andreas@komninos.info