

Editorial

Evaluating affective interactions

Welcome to this special issue of IJHCS, devoted to the exploration of innovative approaches to evaluation of affective systems. We are delighted to present a wide range of papers, culled from a very large number of submissions, which should provide a rich snapshot of the current state of this growing area to the reader.

This collection emerged from a workshop organized at CHI 2005 (see http://www.sics.se/%7Ekia/evaluating_affective_interfaces/), which was in turn driven by an ongoing European Union initiative devoted to evaluation of affective systems (Work Package 9 of the Humaine Network of Excellence, see <http://www.emotion-research.net/> for more information about the broader initiative). The latter has required Höök to wear two hats, both as an active researcher in the area of affective evaluation, and as a curator and gadfly (in her role as head of the Work Package) in helping to shape the overall area of affective evaluation. Isbister initially served as a researcher with the Humaine team, and continues to collaborate with Höök. The 2005 workshop and the present volume reflect our joint efforts at advancing this important area of research and practice, both through particular projects and through the gathering of various efforts and voices toward advancing the field as a whole.

We are particularly proud of the great variance in approaches and stances in this special issue, as they indicate that affective evaluation is by no means a homogenous area. Rather it has turned out to be a crucible for the clarification and discovery of motives and assumptions that help us all to better ground ourselves and to clearly communicate what we are looking for and what we hope to achieve with affective systems.

Evaluation and its role in driving affective interaction efforts

Examining evaluation methods has the significant side-effect of putting the spotlight on problematic issues in Affective Interaction itself. Here are a few key issues that emerged as we assembled the papers in this special issue:

- *Applicability of systems*: Evaluations (both during the design cycle and as a final test) are done in order to check if a system is indeed “worth” the users’ money/attention/use time (in the words of [Gilbert Cockton](#)

(2006)). Thus evaluators help to drive system developers to move beyond figuring out how to do affective input/output or reasoning in an abstract ‘toy problem’ sense, toward incorporating their work into practical, useful application spaces. One can see this in the work of McQuiggan and Lester, who seek to implement effective in-context empathetic behavior from embodied companion agents, and who used a fully functioning commercial game engine (Half-life 2) as the test bed for their work. It can also be seen in Picard and Liu’s pragmatic efforts to utilize affect appropriately to gain better compliance with interrupt-based self-report systems. In both cases, the designers implemented a relatively simple sub-set of larger affective constructs to achieve appreciable results, and McQuiggan and Lester in particular received valuable feedback from end users about whether the emotional model used was helpful/adequate to the context.

- *Underlying models of emotion*: Evaluations in this area have their focus on end-users’ reactions to the systems/applications we build—do they work? This puts the spotlight on what an emotion or an emotional experience really consists of. The papers in this collection reflect a range of perspectives on what emotions are, what in them we can or should measure, and whether it is at all possible to measure them. Contrast the use of the standard accepted positive and negative activation/valence and arousal model of emotion in several of the papers (e.g. McQuiggan and Lester; Hazlett and Benedek; and Smith and Maclean) with the contesting of emotional categorization and emphasis on emergent and polyphonic emotional meaning that appears in the work of Boehner et al.; Sundström et al. and Gaver. Boehner et al. and Isbister et al. in particular contrast universalist, more biologically based models of affect with those that incorporate room for cultural divergence in the practice of emotional expression. Quite a few of the papers explore the integral role of the body in displaying as well as determining emotion (e.g. Sundström et al.; Smith and MacLean; and Isbister et al.), reflecting a larger movement in psychology toward re-blending the body and mind in understanding how emotion intertwines with thought and action. These variations and trends reflect ongoing

debate in the study of affect outside of the design/computation domain, and cannot be overlooked in considering and crafting evaluation techniques.

- *Aims of evaluation tools*: Clarifying outcomes aims shapes the nature of evaluation tools profoundly. For example, the aim of Isbister et al.'s Sensual Evaluation Instrument is to create a rich dialog between designers and users during iteration, rather than to establish a final 'ground truth' assessment of a system's effectiveness, leading to the creation of a free-form self-report measure that avoids words and allows for a wide range of user response patterns. Boehner et al. blend evaluation and innovation as a matter of philosophical stance about the emergent properties of emotion lying between the user and the system, toward richer applications that reveal our assumptions to us in new ways. One could cast Gaver's use of cultural commentators in this same light—providing what he terms a polyphonic assessment of the use conditions that a designer's object presents to those who interact with it. These more interpretive and co-creative approaches can be contrasted with the aims of Hazlett and Benedek or Mandryk and Atkins, who seek to use physiological measures to track real-time emotional responses from users to rapidly shifting experiences of completed applications meant to evoke (or avoid evoking) certain emotions. Both of the latter teams are concerned with creating a useful 'ground truth' picture of emotion that can be used reliably among users of the systems examined. Tähti et al. use commercial product evaluation as a lens through which to evaluate various affective evaluation tools, with an emphasis on the amount of valuable user feedback from each method toward identifying needed modifications toward a successful product launch. In each case, aims become inextricably tied to the selection and meta-evaluation of evaluation means.

As can be seen in this overview, a broad palette of aims, assumptions, and approaches has led to affective evaluation instruments that diverge wildly. One fundamental divide among methods could be categorized as the consideration of how broadly results should be applicable and replicable and in what way. Among those concerned with a ground truth measurement of emotion that is as broadly applicable as possible (e.g. Mandryk and Atkins or Hazlett and Benedek), there is a strong desire for combining subjective and objective measures, for triangulation of measures, and for validation of results through

statistically significant studies and replications. Among those seeking to generate a dialog between designers and end users which does not necessarily require summative and replicable assessments (e.g. Gaver or Sundström et al.), there is rather an emphasis on richness of input, and room for interpretation is seen as an advantage, not as a flaw. The aim is to establish an 'in-the-moment' understanding in the service of particular design goals, which need not extend beyond this context, and along the way (in some cases), to challenge and extend current models and assumptions about emotion and how it contributes to experience of interactive systems.

In both cases, we believe there is a value placed upon rigor—but rigor of a different ilk. If we return to the question of aims, we can see that a designer may require a rich and even divergent and unresolved set of perspectives in order to envision and create the most successful system for engaging and transforming users in their experiences. Whereas someone required to judge a system's efficacy in eliciting particular emotions predictably, requires a reliable instrument of a different sort. The latter rightly ask tough questions of the former about how valid and extensible their measures can truly be, whereas the former can point to limitations and even flaws in fundamental models and assumptions that might otherwise be overlooked. We see these approaches as complementary, and we feel the juxtaposition of stances and aims in this special issue helps to keep all of us 'honest' in our pursuit of powerful and effective affective evaluation methods. We hope this special issue will spark many interesting conversations and innovations to come.

Reference

- Cockton, G., 2006. Designing worth is worth designing. In: Proceedings of the NordiCHI 2006. ACM Press, Oslo, Norway.

Katherine Isbister
Rensselaer Polytechnic Institute,
 110 8th Street, Sage 4208, Troy, NY 12180, USA
 E-mail address: isbisk@rpi.edu

Kristina Höök
Swedish Institute of Computer Science,
 Box 1263, 16429 Kista, Sweden
 E-mail address: kia@sics.se