Flagship Areas in Games User Research

Anders Drachen
AAlborg University
A.C. Meyers Vænge 15
2450 Copenhagen
andersdrachen@gmail.com

Pejman Mirza-Babaei
HCI Games Group, University of
Ontario Institute of Technology
Oshawa, Ontario, Canada L1R 7K4
pejman.m@acm.org

Lennart E. Nacke
HCI Games Group, University of
Ontario Institute of Technology
Oshawa, Ontario, Canada
Lennart.Nacke@acm.org

CHI 2014, Apr 26 - May 01 2014, Toronto, ON, Canada
ACM 978-1-4503-2474-8/14/04.

Abstract
Games User Research (GUR) as a domain of inquiry and as a community is undergoing rapid development in parallel with the innovation occurring in digital games. For example, the re-introduction of virtual reality, the maturing of techniques for behavioral and physiological tracking, attempts for measuring player experience, new contexts of play, the evaluation of learning and ongoing efforts to broaden the target audience of games, all present new opportunities and challenges for GUR work in industry and academia. In this paper, these and other key areas of current GUR work are identified and their potential to shape the future roadmap in the field discussed.

Author Keywords
Games User Research, methods, applications, technologies, contexts, flagship areas

Introduction
When reviewing recent work on Games User Research (GUR), an image emerges, which shows the field to be in rapid development. New technologies, methods, tools and ideas are being presented daily, and GUR today has a well-defined role to play in game development as well as across several academic areas, such as Human-Computer Interaction, Psychology and Data Mining.

GUR has come a long way, as demonstrated by Medlock's [1] review of the history of the field and practice, which also shows how GUR has gained traction within the games industry and in games research. It is, however, also apparent that many of the fundamental challenges in GUR – for example the problem of how to measure Play Experience (PX) – are as important today as they have always been.

GUR as a community also faces a variety of challenges, notably related to knowledge dissemination, -exchange and -preservation. GUR is inherently a multi-disciplinary field, stretching across academia and research. There is, however, an excellent tradition for collaboration across industry and academia, as exemplified by the successful CHI Games User Research Workshop series (CHI-GUR)[1]; the International Game Developer's Associations Special Interest Group on GUR's yearly summits (GUR-SIG)[2], and the promising upcoming CHI PLAY[3] conference series. In a domain with rapid emergence of new knowledge, but a fragmented infrastructure to communicate the knowledge, tracking existing and emerging knowledge is difficult.

In this paper we identify, from the latest GUR work in academia and industry, the areas of current critical interest and development. A thorough but informal review was performed, covering research databases, industry resources and the past three years of the GUR events mentioned above.

[1] http://hcigames.businessandit.uoit.ca/chigur/

[2] http://www.gamesuserresearchsig.org/

[3] http://chiplay.org/

GUR Flagship Areas

From the review, four Flagship Areas were identified: Technologies, Contexts, Methods and Challenges. The term *Flagship Area* is defined as a topic or issue that has either garnered critical interest from professionals, or which is emerging as a new trend with the potential to reshape the roadmap in GUR.

Technologies

In recent years substantial advances have been made in the application of various technologies to academic and industry-based GUR work. These notably fall within the domains of behavioral telemetry, psycho-physiological sensors and virtual reality. In a sense, neither of these technologies are new, but have gained substantially in maturity within the past few years.

BEHAVIORAL TELEMETRY

The application of telemetry tracking to obtain data about in-game and associated behavior from game players has gained substantial traction in both the industry and academic sectors. Behavioral telemetry offers high-resolution, precise data about how users interact with games that traditional GUR methods are hard-pressed to produce. Since the early days of applying behavioral telemetry in GUR, the advent of Free-to-Play (F2P) business models and mobile technologies have seen a rapid adoption of game analytics techniques in the industry. However, telemetry is not always integrated with user testing. A current and future challenge lies in making telemetry analysis available to, notably, small-to-medium sized developers, and developing tools and techniques for integrating telemetry with existing GUR frameworks [2], as well for visualizing behavioral telemetry to make it actionable alongside other user testing results.

Physiological measurements

Different GUR approaches are currently emerging that make different use of physiological data. These approaches are still to be explored in terms of their actual usefulness in game development. Physiological measures may provide important information on emotion that is complementary, or even contradictory, to that provided by self-report or observation measures. They are often regarded as more objective compared to self-reports; they may also provide information on player's feelings that, for some reason (e.g., subtle nature of the responses, repression), are not available to players' conscious awareness. However, to provide context (or ground truth) to sensitive physiological measures, they are usually used in conjunction with other user research methods. Through the use of game logs (or other indexing approaches) that pinpoint exactly when game events were happening we are able to contextualize physiological reactions of players. Overall physiological-based evaluation has been used with two general aims: A) to look for correlations between collected physiological measures (event-based analysis or session-average analysis) and self-report measures as a means of validating the quantitative values captured for *summative evaluation* [2]; B) *Formative evaluation*, to use changes in a player's physiological state as an indexing tool to structure post-session interviews or observations [4].

Virtual reality

Virtual reality (VR) technology has in the past years seen a widespread resurgence in games, especially with the takeover of the VR headset Oculus RIFT by Facebook and Sony's announcement of Project Morpheus at the Game Developers Conference 2014. With the introduction of companies who have started prototyping a new generation of virtual reality technology, – notably head mounted displays – embodied virtual reality games are feasibly within reach. The potential for the game industry in the area is tremendous, however, embodied virtual reality technology comes with a set of associated new challenges for user testing, such as the difficulty of running think-aloud testing with participants wearing head-mounted displays [3]. There is thus a need for new research and development for GUR methods able to handle the unique situation these new immersive technologies provide.

Multi-channel data integration

User research can produce substantial and complex data sets. Several major developers have presented systems for integrating different sources of data, for example behavioral telemetry, screen capture, surveys and face tracking. There is a need for research in this domain, notably with respect to meshing qualitative and quantitative user data.

Contexts

GUR methods for special needs audiences

Providing access to games for all people can be a major challenge – but one that can be addressed in GUR. A main problem for the group of disabled users identifying themselves as Able Gamers is an extended controller support and adherence to multisensory types of feedback (i.e., extending visual feedback to other modalities like audio cues for blind players), since they can often not operate a standard controller or need special gesture input . In addition, younger kids and older adults are becoming increasingly interested in playing games and different devices and often bring

with them special needs regarding the comprehensibility of the interface and controls. As part of an inclusive GUR, we need to be able to provide testing for these special populations.

Free-to-Play: GUR for monetization

F2P and similar business models have necessitated the adoption of business analytics methods to monitor and predict user behavior via the collection of behavioral telemetry. For example, the adoption of techniques such as split-testing (A/B testing or multi-level testing), and the introduction of machine-learning techniques for classifying and predicting player behavior. These goals move outside of the traditional areas of GUR to also cover user spending patterns [2]. There has been some debate as to whether game analytics is also GUR, but, there is at least an overlap between game analytics and GUR. As a new area, there is a substantial room for new theories, methods and techniques that focus on monetization, or attempt to investigate monetization in parallel with user experience and design.

Mobile and other new platforms

The video games industry is experiencing changes in how players interact with games. New platforms provide e.g. intuitive motion-sensitive controllers and less-complex games designed for accessibility to non-gamers. On the other side, the increasing computing power of mobile devices provide increased competition with consoles. These new platforms and ways of interacting with games increases the breath of user research and invites innovative methods to capture and analysis experiences.

Testing outside usability and PX

While the GUR toolset allows testing concepts around usability and player experience, games have left traditional settings and are increasingly used in so-called *gamified* applications, serious games and learning games. Here the goal is often to test how effective a behavioural change or a learning effect is. Novel GUR brain wave methods are promising for getting insights into learning and decision-making in games. However, GUR still needs to establish a standardised toolset and methods for this separate market where the goal of the game is not to play, but to learn something or motivate a behaviour.

Methods

Play heuristics

Although heuristic evaluation promises to be a low-cost usability evaluation method, it suffers significantly with problems concerning evaluators' subjective interpretations . To answer this limitation, researchers have aimed to develop a more specific set of heuristic. For example, to fit a certain game platformor critic proofing approach that takes into account a problem's frequency, impact, persistence and a game's genre.

Automated testing: lab-based testing, data and analysis done in an automated fashion

Different testing scenarios bring different experimental requirements with them. For certain GUR testing approaches, a lab-based environment is paramount to rule out any confounding factors (e.g., environmental influences that can affect volatile physiological measures). For other approaches, automated data collection and analysis is preferable. Automation of in-game data collection and triggering is always better when facing large amounts of data being collected. The

rule here seems to be that higher density data is extremely hard to analyse and cross-correlate with other GUR measures manually, so that automated markup and processing of the data is desirable. Protocols are currently being established, but so far every researchers and many industrial GURs are building their own solutions, which is justified as there might be specific needs for a particular team.

Remote testing

Remote testing in the context of GUR is simply the application of techniques for obtaining information about user behavior and user experience over a distance. Remote testing has gained traction recently thanks to the introduction of behavioral telemetry tracking and associated analytics [2], but remains a relatively undeveloped area of GUR. For example, remote surveying and video capture provide the ability to capture user information at a distance, or in contexts where laboratory-based methods are not ideal. The potential for new tools and methods in this area is substantial, opening up larger sample sizes and cheap alternatives to laboratory testing. However, there is a costs-benefit balance, as remote testing requires the flexibility to work in conditions, where there is less control than in the laboratory.

Low-cost testing

Although conducting 'formal' GUR studies as part of video games development cycle has become more popular over the last decade, but apart from cases from large developers and publishers there are not many reports if smaller to midsize studies have fully applied these measures. Barriers toward adapting GUR studies, (such as that personnel need to be trained extensively in interpreting and correlating multiple data sources) may seem out of scope for small to midsize development companies. One step towards making GUR studies more available for larger game developer community is to develop low-cost testing methods (methodological optimizations suitable for small scale GUR studies). This would have a direct impact on including GUR in the development process and on return on investment.

Challenges

Integrating GUR in iterative development

A substantial number of presentations and discussions in the past few years at the CHI GUR and GUR SIG events, as well as larger events such as the Game Developers Conference, have focused on challenge of integrating GUR in iterative game development [2].

This central challenge deals with how to obtain valid and generalizable information from playtesters in a way that is financially viable, flexible to a variety of play contexts and goals (e.g., feedback on gameplay, art/graphics, mechanics, narrative), which delivers actionable insights, and which is fast so that it can be integrated in a rapid iterative development process. This is not a new challenge, but rather one of the fundamental issues that underlies GUR work.

The vast majority of the work is being driven by large studios, with the capacity to build in-house test labs and hire experts. Less work has been focused on developing solutions for small-medium sized developers. Irrespective, the introduction of new technologies, business models and play contexts mean that the practical challenge of user-testing games is getting progressively more complex.

MEASURING PLAYER EXPERIENCE

One of the largest challenges that the GUR community is currently facing is the development of good player experience models [5] that can be used in game design and GUR. The challenge of measuring player experience is that some aspects like enjoyment cannot be measured as solidly in more objective measures as they can be in self-report measures (e.g., PENS). The major challenge for GUR remains to find a good combination of measures that can identify a valuable player experience holistically, efficiently and cheaply.

Conclusion

A number of new technologies have emerged and gained maturity in recent years, all of which have carried with them new challenges and opportunities for user research. At the same time, the target audience for games has diversified as have the contexts of play, which means that user testing has to accommodate these scenarios. It is worth noting that the Flagship Areas identified above do not exist independently. For example, the emergence of new mobile platforms, expansion of the target audience for games, combined with free-to-play business models and behavioral tracking, creates a need for GUR techniques to handle large-scale user data, combine in-house testing with remotely captured data, and integrate all of this with the iterative rhythm of game development.

While the Flagship Areas discussed here are not intended as a definitive list, they form a starting point for discussion about the current status and future of the GUR domain.

Acknowledgements

Thank you to GRAND NCE, SSHRC IMMERSe, and NSERC for funding this research.

References

[1] Medlock, M. History of Video Games User Research. Presentation at the IGDA GUR Summit (San Francisco, CA), 2014.

[2] In Seif El-Nasr, M., Drachen, A., & Canossa, A. (Ed.), In Game Analytics (pp. 585-619). Springer London. doi:10.1007/978-1-4471-4769-5_26

[3] Guardini, P. GUR, Put your helmet on: user testing virtual worlds. Presentation at the IGDA GUR Summit (San Francisco, CA), 2014.

[4] Mirza-Babaei, P., Nacke, L. E., Gregory, J., Collins, N., & Fitzpatrick, G.. (2013). How does it play better? Exploring user testing and biometric storyboards in games user research. Proceedings of CHI 2013, New York, NY, USA. doi: 10.1145/2470654.2466200

[5] Engl, S., Lennart E. Nacke, Contextual influences on mobile player experience – A game user experience model, Entertainment Computing, Volume 4, Issue 1, February 2013, Pages 83-91, ISSN 1875-9521, doi: 10.1016/j.entcom.2012.06.001.