Playing Around the Eye Tracker: A Serious Game Based Dataset

Michael Riegler[1], Ragnhild Eg[1], Lilian Calvet[1], Mathias Lux[2], Pål Halvorsen[1], Carsten Griwodz[1]
[1]Media Performance Group, Simula Research Laboratory & University of Oslo, Norway
[2]Institute for Information Technology, University of Klagenfurt, Austria
{michael, rage, paalh, lcalvet, griff}@simula.no, mlux@itec.aau.at

Abstract

This work applies crowdsourcing and gamification approaches to the study of human visual perception and attention. With the presented dataset, we wish to contribute raw data on the salience of image segments. The data collection takes place in the designed game, where players are tasked with guessing the content of a gradually uncovered image. Because the image is uncovered tile-by-tile, the game mechanics allow us to collect information on the image segments that are most important to identifying the image content. The dataset can be applied to both computer vision and image retrieval algorithms, aiming to build on the current understanding of human visual perception and attention. Moreover, the end objective is to test the game as a potential substitute to professional eye tracking systems.

1 Introduction

In the ongoing quest to understand how humans think, perceive, and behave, human computation and related fields contribute with new methodologies and models that can shed more light on the complex workings of the human mind. Researchers make use of human computation to train machines, such as in semi-supervised learning, but also to collect data on tasks that can only be completed by humans. Crowdsourcing is often applied to this kind of data collection, alleviating the burden of running experiments, but at the same time introducing a few methodological concerns. Moving experiments and user studies out of the restricted environment of a laboratory means surrendering control over the test situation. Fortunately, there are means to compensate for the lack of control. Crowdsourcing makes it far easier and less time-consuming to collect data from a large number of people, improving both the internal validity and the generalisability of results [KCS08, NR10]. Another concern is linked to the motivation of crowdsourcing workers and their willingness to adhere strictly to the task at hand. However, this concern can be addressed through the design of the experimental task [KCS08]. Gamification is a fairly recent development in crowdsourcing. Games with a purpose (GWAP) provide an entertaining arena for participants, aiming to enhance player motivation and improve performance. Hence,

In: F. Hopfgartner, G. Kazai, U. Kruschwitz, and M. Meder (eds.): Proceedings of the GamifIR'15 Workshop, Vienna, Austria, 29-March-2015, published at http://ceur-ws.org

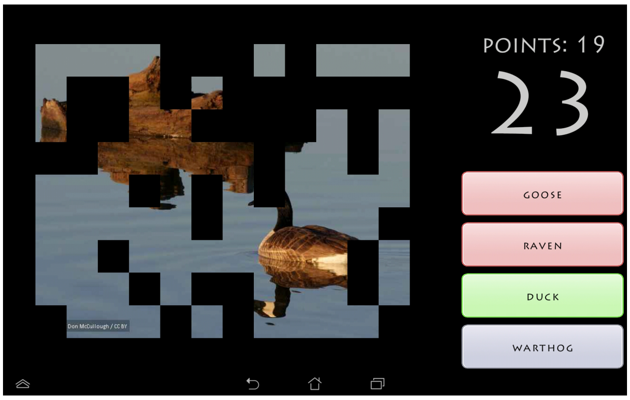

Figure 1: The standard difficulty mode of the game, shown as an example round with two incorrect attempts (red buttons) and the final correct response (green button). Accumulated points and remaining time are presented in the upper right corner.

through a well-designed GWAP, researchers retain the benefit of reaching out to a large pool of participants, while increasing the likelihood of obtaining more reliable data on problems only humans can tackle. With this approach, researchers have succeeded in turning annotation tasks into enjoyable activities [VAD08], along with a range of other repetitive tasks, ranging from information retrieval to security and exertion issues [JKM+11, BMI14, PMRS14, MBB10]. Furthermore, some games have been designed to tap directly into processes that involve human visual perception and attention. For instance, Peekaboom is a two-player game where one player is asked to guess the content of the image that the other player is gradually revealing [VLB06]; another approach presents players with a short video that is subsequently masked by a character chart [RGSZM12]. In both games, the collected data is used to shed light on where and what people will look at in an image or a video.

Our gamification approach is similarly motivated by questions on how people regard and recognise a depicted object of interest. Human visual attention is typically studied using eye tracking paradigms, with equipment that can map the movements and fixations of the pupil, for instance across a presentation on a screen (e.g., [HRM11]). Researchers have used this technology for decades to study how the eyes move during reading, and this body of work has established important insights on the processing of written information [Ray09]. Eye tracking methods are also used to explore where and how people look when taking in a scene [CMH09], when performing a visual search task [BB09], or when looking at the face of a someone talking [BPM08]. Humans are in fact quite adept at recognising people, objects and animals, even with fairly degraded global features [McMM11]. Although human visual perception is facilitated by higher-level cognitive mechanisms, such as prototypes stored in long-term memory, the visual system relies on attended low-level features that may be unique to a particular animal or object [McMM11]. Furthermore, attention is easily captured by visually distinct or unexpected elements within a scene [BB09]. The limited number of studies into salient regions and features involved in the identification of objects and animals could very well be connected to the time needed for such an undertaking. Running dozens of individual eye-tracking sessions with hundreds of images seems a daunting task, not to mention an expensive one (see for instance [MCBT06, JEDT09, BJD+]). With this in mind, we planned our serious game as a time-efficient and economical alternative to traditional eye-tracking paradigms.

Inspired by research on human vision and attention, computer vision scientists work to overcome the problems of computational complexity in order to replicate the mechanisms of human perception. By building such systems, researchers in this field aim to solve problems related to object recognition and scene interpretation, as well as other related challenges. When addressing human visual attention, one term becomes particularly prominent in both cognitive psychology and computer vision. In psychology, *visual saliency* can be determined by the low-level features that affect where people move their gaze, such as contrast, colour, intensity, brightness, and spatial frequency [Ray09]. Similar definitions have been proposed by the computer vision community. *Saliency*, as defined by Borji and colleagues [BI13], *"intuitively characterizes some parts of a scene— which could be objects or regions — that appear to an observer to stand out relative to their neighboring parts."*. Humans are able to identify salient areas in their visual fields with surprising speed and accuracy before performing actual recognition. This remains a critical task in computer vision. To assist in this endeavour, we wish to supply the multimedia and computer vision communities with a dataset that can be useful:

- As input data for machine learning algorithms aiming to detect salient objects/regions.
- As input data for scalable image understanding systems: feeding a few salient regions into thousands of object classifiers (e.g., [NKRP10]) without running thousands of expensive object detectors across the entire image.
- To evaluate computational methods of salience (such as [BJD+, JEDT09]).

Our game is designed to gradually reveal parts of an animal picture (although the game can easily be adapted to other types of images) and the player's task is to identify the animal as quickly and as accurately as possible. Because the various elements are revealed in a random pattern, the game makes it possible to analyse response patterns and explore which regions are most vital to the recognition of the animal. Furthermore, the crowdsourcing arena enables comprehensive data collection, securing sufficient data for the analyses of the separate images.

In the provided dataset, we have collected image unveiling patterns and the related subjective responses. Through the design of our crowdsourcing study we have created a novel single-player game that entertains and engages participants, aiming to increase motivation and divert attention and awareness away from the underlying research question. Along with the game and the stimulus material, we provide data from our first rounds of experimentation. With this material, we wish to:

- Make the Mobile Picture Guess game publicly available as a low-threshold experiment set-up.
- Provide an openly available dataset for investigations into human visual attention and salient image features.
- Provide a dataset that can be compared with results collected from an eye tracker, and in turn explore the feasibility of our approach as an alternative to these costly systems.
- Finalise our investigations by establishing salient features for individual animal images, hopefully building on the current understanding of human visual perception.

The planning, design, launch and analysis of our serious game progressed over several stages that we describe in the paper, beginning with the technical design and the data collection. We then include details on our dataset and outline the experiment we conducted to highlight the application of the game including a preliminary analysis. Finally we draw our concluding remarks.

2 Data Collection and Game Design

2.1 The Game

The game, designed to entertain while collecting data, is called Mobile Picture Guess [RELS14]. It involves a puzzle, a gradually revealed image, that must be solved before time runs out. The way to solve the puzzle is to guess the content of the image by choosing the correct option out of the four presented; in the current set-up, all images portray an animal, thus all response options provide an animal name. Based on feedback from initial user tests, parts of the game were modified over multiple permutations. The end result is a game that is fun to play and that gathers data without intrusion.

2.2 Technical Details

The full data collection system consists of two parts. One part is the game running on Android devices, the other is the back end server solution. The main development platform of the game is Java, using libGDX[1] and the Android library[2]. LibGDX is an open source framework for cross platform game development. It provides an easy way to create 2D interactive programs based on OpenGL on MacOS, Windows, Linux, Android, iOS and Web applications. While it is mainly targeted at Android platforms, Mobile Picture Guess can easily be adapted to other platforms. We decided to use the Google Play functionality to distribute the game to a large amount of players.

The back end server is an Apache web server[3] hosted by our lab. It also runs a MySQL server. HTTP requests over a PHP based script are used for the communication between the server and the game. To provide maximal security for the player and the data, we employ several techniques to avoid SQL injection, along with strong data encryption. The server retrieves the data from the game in a JSON file format and stores it in the MySQL database. To make this possible, the player has to be on-line while playing the game. The information is stored after each image's revelation, in order to avoid its loss through cancelled games or interrupted internet connections.

2.3 Gameplay

With the overall aim to collect perceptual information, the game task needs to capture the full attention of the player, necessitating a single-player design. The game mechanics involve: a puzzle to be solved, play against time, adaptive difficulty for skilled players, and a scoring expressed by points. A player starts a new game with a contingent of time, and for each round the player is presented with a new image to guess the content of. One completed game consists of as many rounds as the player can complete in the given amount of time. At the beginning of each round, the image is completely obscured by black mosaics. These black tiles commence to disappear in a random pattern as time counts down, illustrated in Figure 1. Thus, the image surface becomes gradually more visible as the mosaic is lifted. The longer the round runs, the easier it becomes to guess the image content. If players cannot complete the task before time runs out and the image is completely unveiled, they receive no points and the game skips ahead to the next round.

In order to provide a response on the image content, the player chooses one of the four alternatives presented as buttons on the right side of the screen. Three of the buttons display incorrect answers and one of them holds the right one. The player clicks on the option presumed to be correct, then receives immediate feedback. Incorrect answers will turn the selected option red, whereas correct answers will yield a green button (Figure 1). With the right answer provided, the picture is fully revealed and the player receives points and additional seconds of playtime. The number of points is based on how much of the picture remains concealed, and thus depends on the swiftness of the response. Furthermore, wrong answers result in loss of playtime and this loss increase steadily with repeated

[1] http://libgdx.badlogicgames.com/

[2] http://developer.android.com/develop/index.html

[3] http://www.apache.org/

incorrect responses. This accumulated loss penalise attempts at choosing all options rapidly without focusing on the image. We implemented this reward and penalty system in order to motivate players to play as quickly and as accurately as they could. To ease the task of learning the game rules, the game becomes more difficult over time. The easy mode at game start involves a high rate of tiles revealed, meaning that tiles disappear more quickly. At the successful completion of the initial rounds, the rate of uncovery goes down. Moreover, a transformation is applied to the picture to make it harder to guess the content, exemplified in Figure 3. This transformation is completed by flipping the image 180 degrees, changing the colour randomly, or changing the colour to greyscale. The uncovery reduction and the transformation are applied solely to make the game harder, and consequently more interesting for players who may want to play additional rounds. For variation and less monotony, we also implemented a mini-game. The game is simple, but requires some dexterity from the player. The task is to reveal a concealed image by sliding a finger across the screen to remove the black squares, as portrayed in Figure 2. If the player succeeds in uncovering the entire image, they receive additional time for the main game. Thus, the mini-game serves two purposes. On one hand, it introduces a new task to distract the player from the potential repetitiveness of the main game, hopefully improving the quality of experience. On the other hand, it works as an aid to improve performance on the original task. The mini-game is presented after five image-guessing rounds; if it is successfully completed, three bonus seconds are added to the play-time. Please note that data from the mini game is not collected for the dataset. The game continues until time runs out, at which point the player is presented with their final score. The game then returns to the start screen, where the overall high-score is listed and the player can choose whether to play another game or to end the session.

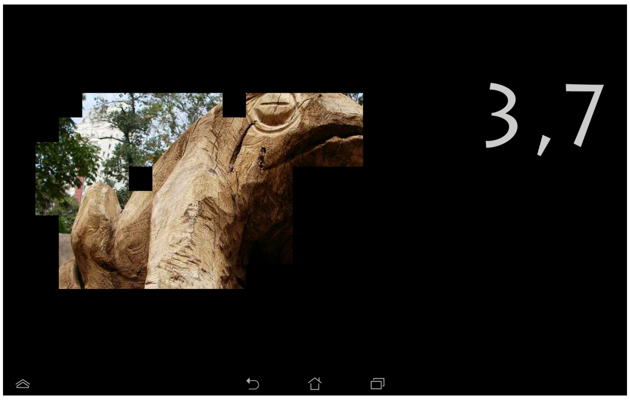

Figure 2: An example image from the mini-game in Mobile Picture Guess. The image on the left starts out as a black rectangle that is uncovered by sliding a finger across the tiles, on the right is the remaining time.

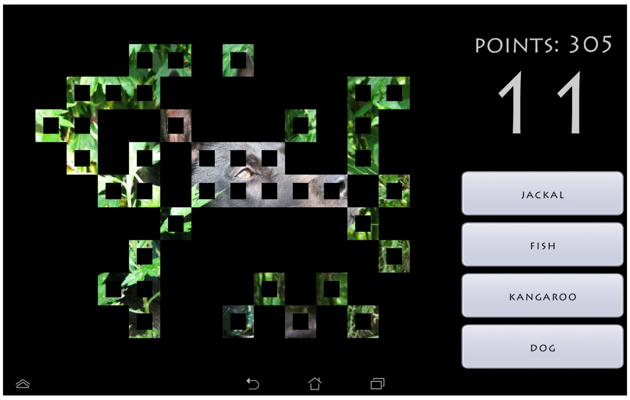

Figure 3: Screenshot of a round from the difficult game mode, where less information is revealed when the tiles are lifted. This mode is introduced further into the game, after a player has successfully completed several rounds.

2.4 Human Intelligence Task

Because a new game in the app store is easily overlooked, we also made use of crowdsourcing in our recruitment of players. As crowdsourcing platform we used Microworkers[4]. In the HIT we asked the workers to download the game and play it. We added a token system to make sure the workers dedicated both time and effort to the game, and each worker had to report two tokens per task. Initially, one token required 2000 game points, but because of the observed difficulty in reaching this mark, we reduced it to 1500 points. Feedback from workers suggest that the game was well-received and enjoyable to play. We ran the HIT for one week; based on recommendations by Microworkers we paid workers 0.80 Euros per HIT. In total we spent 100 Euros on the whole experiment, including the fee for the Microworkers platform. Additional information collected about the HIT and the games played are presented in Table 1. Sadly, our game did not run properly on some of the older Android devices; this required us to investigate our dataset manually and exclude scores collected from these devices. However, the workers were not affected by this issue.

[4] https://microworkers.com/

3 Dataset Description

The publicly available dataset contains 200 images, in addition to the SQL database file with the collected player information. An overview of all dataset components is illustrated in Figure 4.

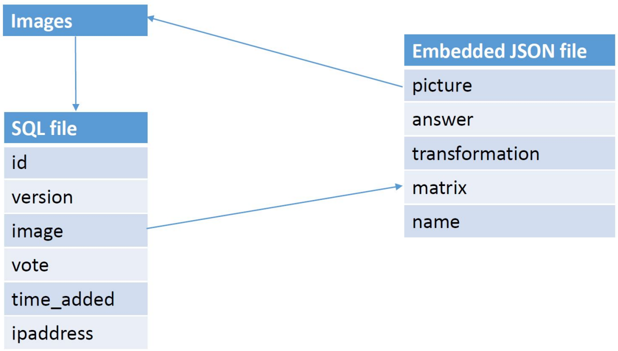

Figure 4: Overview of dataset components.

Completed games	13.861
Submitted HITs	111
Unique workers	302
Unique players	352
Average games played	5
Average game-time	10 minutes

Table 1: General statistics about the dataset.

Selected images. To create our image dataset, we first settled on a list of 124 animals so that each image could be easily distinguished and described by a single label, such as albatross, alligator, and alpaca. Next, we used these terms to query images on Flickr, collecting images categorised as *free to use* or published under a Creative Commons attribution license[5]. We made sure to select visually appealing scenes by ranking the queried images according to Flickr's interestingness score and then keeping the highest ranked 25 images for each term. The resulting dataset, with more than 3000 images, was further reduced to 200 by removing all manipulated photos and all images that did not clearly display the animal of interest. For each image presentation, we added three random terms to the correct label, yielding the four response options.

Statistics. By releasing the game on Google Play Store, we could easily keep track of the application and the data collection. It also allowed us to derive statistics on the games played, these are summarised in Table 1. As noted before [RELS14], several workers continued to play without payment, after completing their HITs. This provided us with additional data; more importantly, it further established the entertainment value of the game.

Database. All image metadata are stored in an SQL database, which we have made publicly available for download at `http://goo.gl/CL24aV`. Along with the database we include the code to calculate region saliency. The database file consists of six fields; the players' responses are contained in the 'vote' field, which is further divided into five sub-fields. Details about information stored in the respective fields are included in Table 2.

[5]A license text was overlaid Creative Commons images.

ID	Unique ID for the played game
Version	Game version
Image	Image file name
Time added	Time of data submission for the completed game
IP address	Encrypted and secure version of the player's IP address
Vote	Detailed information about the game played (in JSON format)
- Picture	Picture name
- Answer	Correct image label
- Name	Unique device ID
- Matrix	Number of tiles removed
- Transformation	Applied image transformation

Table 2: Description of database fields.

4 Application of the Dataset

With the outlined dataset, we aim to provide information about images and responses presented and collected in a puzzle game. In the game itself, players are tasked with guessing the content of an image that is gradually revealed. The saliency of each image region corresponds to its importance in identifying the content. Specifically, the saliency of an image segment is determined by the number of times the tile was uncovered prior to a correct response, aggregated across all players and divided by the number of times the image was presented in a game. The saliency scores can be mapped out across the images, yielding visual heatmaps. Examples of the aggregated heatmaps are presented in Figure 5, where the density of the colour red is inversely related to the saliency of the tile. Additionally, the saliency value is provided in the lower left corner of each tile.

While the dataset can provide insights directly from the salient regions, which we have explored in [RELS14], it can also be a useful addition to several related research areas. As mentioned, computer vision scientists could use the data to feed into learning algorithms for saliency detection or into image understanding systems, or the data can be applied in evaluations of existing computational methods. Moreover, the dataset can serve as a ground for comparison to eye-tracking paradigms. This is also the next step in our studies into human perception of image regions.

5 Conclusion

In this paper we have presented a dataset that can be applied to (i) improve understanding of human vi-

Figure 5: Example images with the aggregated experiment results, with overlaid saliency heatmaps. The red colour density is inversely related to saliency and the exact value is provided in the bottom left hand corner. The images depict a) a bison, b) a gerbil, c) a sparrow, d) another bison, e) a dolphin, and f) a lynx.

sual perception and attention for image scenes, and (ii) lead a new direction on how information can be collected more efficiently, providing an alternative to expensive and time-consuming eye-tracking studies. Furthermore, we have described the game design and the data collection and provided an overview of potential application areas.

We plan to extend on this work by comparing the saliency scores from the game with saliency data collected using eye-tracking techniques. Through this endeavour, we will be able to explore whether our method yields comparable results and can be used as an alternative to traditional eye-tracking studies. Furthermore, future works should include images that depict different types of scenes and objects, hence extending the exisitng dataset.

6 Acknowledgements

This work is partly funded by the FRINATEK project "EONS" (#231687) and the iAD Centre for Research-based Innovation (#174867) by the Norwegian Research Council and the Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant KWF-20214/25557/37319.

References

[BB09] James R Brockmole and Walter R Boot. Should I stay or should I go? Attentional disengagement from visually unique and unexpected items at fixation. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3):808–815, June 2009.

[BI13] Ali Borji and Laurent Itti. State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207, 2013.

[BJD+] Zoya Bylinskii, Tilke Judd, Frédo Durand, Aude Oliva, and Antonio Torralba. MIT saliency benchmark. http://saliency.mit.edu/.

[BMI14] Markus Brenner, Navid Mirza, and Ebroul Izquierdo. People recognition using gamified ambiguous feedback. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, pages 22–26, Amsterdam, 2014. ACM.

[BPM08] Julie N Buchan, Martin Paré, and Kevin G Munhall. The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research*, 1242:162–171, November 2008.

[CMH09] Monica S Castelhano, Michael L Mack, and John M Henderson. Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):1–15, 2009.

[HRM11] Falk Huettig, Joost Rommers, and Antje S Meyer. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2):151–171, June 2011.

[JEDT09] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2106–2113, Kyoto, 2009.

[JKM+11] Craig Jordan, Matt Knapp, Dan Mitchell, Mark Claypool, and Kathi Fisler. Countermeasures: a game for teaching computer security. In *Proceedings of the 10th Annual Workshop on Network and Systems Support for Games*, page 7, Ottawa, 2011. IEEE Press.

[KCS08] Aniket Kittur, Ed H Chi, and Bongwon Suh. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–456, Florence, 2008.

[MBB10] Florian ”Floyd” Mueller and Nadia Bianchi-Berthouze. Evaluating exertion games. *Evaluating User Experience in Games*, pages 187–207, 2010.

[MCBT06] Olivier Le Meur, Patrick Le Callet, Dominique Barba, and Dominique Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):802–817, 2006.

[McMM11] Vasile V Moca, Ioana Ţincaş, Lucia Melloni, and Raul C Mureşan. Visual exploration and object recognition by lattice deformation. *PloS One*, 6(7):e22831, January 2011.

[NKRP10] Vidhya Navalpakkam, Christof Koch, Antonio Rangel, and Pietro Perona. Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences*, 107(11):5232–5237, 2010.

[NR10] Stefanie Nowak and Stefan Rüger. How reliable are annotations via crowdsourcing? A study about inter-annotator agreement for multi-label image annotation. In *MIR ’10 - Proceedings of the International Conference on Multimedia Information Retrieval*, pages 557–566, Philadelphia, 2010.

[PMRS14] Dinesh Pothineni, Pratik Mishra, Aadil Rasheed, and Deepak Sundararajan. Incentive design to mould online behavior: a game mechanics perspective. In *Proceedings of the First International Workshop on Gamification for Information Retrieval*, pages 27–32, Amsterdam, 2014. ACM.

[Ray09] Keith Rayner. Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8):1457–1506, August 2009.

[RELS14] Michael Riegler, Ragnhild Eg, Mathias Lux, and Makrus Schicho. Mobile picture guess: A crowdsourced serious game for simulating human perception. In *Proceedings of the SoHuman Workshop 2014*, Barcelona, 2014. Springer.

[RGSZM12] Dmitry Rudoy, Dan B Goldman, Eli Shechtman, and Lihi Zelnik-Manor. Crowdsourcing gaze data collection. In *Proceedings of the Conference on Collective Intelligence*, Cambridge, MA, 2012.

[VAD08] Luis Von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.

[VLB06] Luis Von Ahn, Ruoran Liu, and Manuel Blum. Peekaboom: A game for locating objects in images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 55–64, Montréal, 2006.