

From Cookies to Puppies to Athletes: Designing a Visual Audience Voting System

Kirstie Hawkey, Melanie Kellar, Bonnie MacKay, Karen Parker, Derek Reilly

Faculty of Computer Science, Dalhousie University
6050 University Avenue
Halifax, Nova Scotia, B3H 1W5, Canada
{ hawkey, melanie, bmackay, parker, reilly } @cs.dal.ca

Abstract

We describe the design process taken and prototype developed for a system supporting audience voting at the Summer Olympics. In our examination of issues related to the expression and capture of audience votes, we utilized informal focus groups, interviews with experts, and user studies comparing voting formats. Our proposed solution involves displaying results on large screens. This provides the context for visual audience interaction through the use of ranking booklets. The audience input is determined through image analysis of the audience.

Categories & Subject Descriptors: H.5.3 [Information Interfaces and Presentations (e.g. HCI)]: Group and Organization Interfaces.

General Terms: Design, Human Factors.

Keywords: Audience participation, voting, design process, large screen display, image analysis.

INTRODUCTION

Our task for the CHI Student Competition was to design an audience participation system to be piloted at the diving and gymnastics competitions at the 2004 Summer Olympics. The audience must cast their votes in real time and the system must be flexible enough for use in different sports. The system must also be usable by a diverse international audience, tamper resistant, reliable, and cost effective.

This paper describes the development of our solution to the problem. We begin by providing related work in the field of audience participation systems. We next describe our design process, including user studies that guided our solution. We then discuss issues relating to our voting scheme and arising from limitations of current technology. Our proposed solution is then presented including the results of a user study evaluating our approach and the work necessary to advance our prototype.

BACKGROUND

Audience participation is becoming popular across a variety of entertainment forums. Audiences can be more involved

in an event and express their knowledge or opinions about an event or the competitors. On television, at-home viewers of *American Idol* determine the winner by voting via telephone or text message; while the in-studio audience of *Who Wants to be a Millionaire* can vote for the correct answer in the lifeline "ask the audience".

At SIGGRAPH '91, Loren and Rachel Carpenter presented an audience participation system [1] that is still in use today (www.cinematrix.com). The system can detect whether audience members are holding up the green or red side of a reflective paddle. The paddles have been used for a variety of activities, from simple voting to a large game of Pong. Maynes-Aminzade et al. [4] expanded on this work testing three other audience participation techniques. The audience controlled a game of Pong by leaning to the left or right. Also, shadow tracking was used to bat a beach ball around and laser pointers controlled a collaborative paint program. A novel input technique developed by Deleeuw [2] could be applied to audience participation scenarios. Instead of traditional input devices, the input signal consists of props such as paddles or colored gloves worn by members of large audiences. These props are recorded using a video camera and then detected and tracked within the image.

Inselberg [3] describes an interactive audience participation system where each user is provided with an interactive device, preferably wireless. The device includes a graphical user interface and an audio channel for expert commentary and play-by-play information. During the event, spectators can use the device to answer trivia questions or vote. Their feedback is then processed and broadcast to the entire audience through the individual devices or a large screen in the stadium. This system also displays sponsor messages.

DESIGN PROCESS

Our first goal was to identify design constraints. We collected information such as scoring details, size of venues, audience makeup, length of events, and number of competitors. Ongoing round table discussions were critical to the evolution of our prototype. During the meetings, we synthesized information gathered through various methods, including consultations with experts and experimentation, to refine design requirements. Initial brainstorming focused on the nature of evaluation itself, regardless of interface,

with tentative ideas about design. Our experts included informal focus groups of HCI and information retrieval students and faculty. We also spoke with experts in both gymnastics and diving to help us understand the nature and format of the competitions and their typical audience.

We conducted two experiments with students, faculty, and staff. One compared the evaluation techniques of ranking and rating in a group context. The second further explored group evaluation with ranking, and tested the use of a preliminary flipbook design for voting booklets. In later phases, we focused on prototype development, researching the technical feasibility of the design and interface aspects. We then conducted a wizard-of-Oz test of the prototype in context, and evaluated the results.

User Study: Ranking vs. Scoring

The scoring schemes used to judge Olympic events such as diving and gymnastics can be complex and require expert knowledge of the sport. For example, women's artistic gymnastics are scored out of nine possible points with a bonus point available for special combinations or difficult elements. Deductions of varying values occur if required elements are not completed or if faults are noted. Judges use a shorthand system to record the routine as it occurs so that they can quickly determine their score.

Olympic audience members have varying knowledge of the sporting events, limited resources to use in judging the athletes, and often a blatant country bias. Using the regular scoring scheme for a sport could be too slow or complicated to allow full audience participation. We examined the voting paradigms of ranking and scoring with a user study.

To compare ranking and scoring, we conducted a cookie-tasting contest, inviting 21 participants to judge five different chocolate chip cookies. Two acted as our "expert" judges, representing the real judges in an Olympic sporting event, and scored the cookies using a detailed scoring scheme. Their scores were recorded for comparison with the audience's results. The other 19 participants were asked to be our audience: nine used the ranking method of judging, and ten used the scoring method. The entire audience tasted each cookie at the same time. Each audience member then assigned the cookie either a rank or a score.

Ranking participants ranked each cookie in relation to the other cookies that had been tasted thus far. Scoring participants gave each cookie a score between 1 and 10. Although they could use decimal places, most chose not to. After each cookie, the results of both the ranking and scoring were totaled and the average scores and ranks were written on a large whiteboard, along with the results from the two "expert" judges.

After the contest was finished, participants completed a brief questionnaire. The majority (see Table 1) of both groups indicated their method allowed them to easily judge the cookies. However, there was a significant impact of

	Did your method allow you to easily judge the cookies?		Would you have preferred to use the other method?		
	Yes	No	Yes	No	Don't Care
Rankers	7	2	0	6	3
Scorers	7	3	7	2	1

Table 1. Ranking vs. Scoring: ease of use and preference

voting type on whether users would prefer to use the other method ($\chi^2(1, N=18) = 9.975, p<.01$) with seven of the scorers preferring to rank but no rankers preferring to score.

The results suggest that ranking may be preferred over scoring for audience voting; however, in our study, the scoring method was not particularly complex and the audience was as "expert" as the judges. We suspect that the preference for ranking will hold when members of the audience lack sufficient knowledge of the sport being judged to understand the scoring regulations well enough to consistently and accurately apply them.

User Study: Top 5 vs. Top 10

We next examined whether there was a preference by users for ranking a particular number of competitors through a mock similarity competition of 31 pictures of people who look like their dogs. This was a between subjects study where one group of ten users ranked the *Top 5* competitors and, later, a second group of nine users ranked the *Top 10* competitors. Users could rank a participant in the *Top X* or move the competitor out of the running by ranking $>X$. Each user was given a voting booklet prototype. After each picture was viewed, users selected and held up a ranking card from their booklet. The median of the scores became the ranking for the current participant. The resulting *Top 5* or *Top 10* competitor's pictures were placed on the whiteboard in their ranked order along with the scores of two judges. After the competition, users completed a short questionnaire.

There was no clear consensus as to the preferred number of competitors to rank. Interestingly, the majority of users (7/10) who ranked the *Top 5* felt they would prefer to rank the *Top 10* and vice versa (6/9). One user who ranked the *Top 10* noted that "5 is not enough flexibility", while another *Top 10* user found it was "too hard to keep all [the competitors] in memory" and preferred to rank only the *Top 5*. Another user noted that since only 3 competitors in the Olympics actually medal, ranking the *Top 5* was best. The users also indicated on a 5-point Likert scale how easy it was to *select* and *display* their ranking using the scoring booklets (see Table 2). Users indicated that it was Somewhat Easy or Easy to *select* ($t(18) = 1.68, p=.055$) and *display* ($t(18) = 3.98, p<0.001$) their rankings. Flaws in the quality of our prototype (defective binder rings and staples) were mentioned by the users who had some difficulty.

User Group	Easy	Somewhat Easy	Medium	Somewhat Hard	Difficult
Ease of Selecting					
Ranked 5	4	3	2	1	0
Ranked 10	2	4	3	0	0
Total	6	7	5	1	0
Ease of Displaying					
Ranked 5	6	2	1	1	0
Ranked 10	5	3	0	1	0
Total	11	5	1	2	0

Table 2. Perceived ease of selecting and displaying ranking cards

We observed that users ranking the *Top 5* competitors were more orderly, generally less confused, and faster selecting their cards and scoring. In contrast, users ranking the *Top 10* competitors seemed more indecisive, took longer and were influenced more by the general audience ranking.

ISSUES

Audience Method of Judging

The choice of a small, collective, audience ranking scheme is the most important aspect of our design. It influences all aspects of our interface design, and defines the very nature of the activity being supported. Therefore, we needed to properly understand the mechanics of our scoring design in its intended context. We, therefore, conducted ongoing critical analyses of the technique among ourselves, with experts and with experimental participants.

Ranking is not the same as judging. Spectators may want to score in a fashion similar to the actual judging. Our gymnastic expert said that many in the audience have sophisticated knowledge of the events and scoring. In addition, a score based on specific criteria may remove some of the “crowd pleaser” influence and membership bias. However, authentic judging is more time-consuming and may detract from the fun of the event, particularly for those not as familiar with the scoring criteria. A simple ranking system may allow greater audience participation and a more visceral, immediate collaborative voting experience. Our gymnastics expert indicated that audience members tend to see the “big picture” more than minutiae, and that this is a valuable evaluation that differs from official judging. Group-ranking with a small visible top-ranked group is arguably a better fit to audience participation than the formal scoring performed by judges.

We debated whether the audience should be ranking each competitor against their *own* previous rankings or against the *aggregate* audience rankings. While audience members may prefer to rank against their own previous rankings, this requires providing some method for each audience member to track their previous rankings. This approach also

presents problems in providing a combined audience result unless each audience member submits a fully ranked list at the end after ranking each athlete. Therefore, we settled on the simpler solution of an aggregate set of standings for the audience to rank each competitor against.

Ranking requires audience participation and concentration throughout an entire event. Ranking new performances is difficult if spectators miss some athletes or forget details of previous performances, a particular concern in gymnastics with many competitors and simultaneous events. Therefore, some spectators may be ranking the current athlete against others whose performances they did not see. The ability to also see the judges' scores and the audience consensus of athletes missed, in relation to other athletes the spectator has seen, may mitigate this problem.

As there was no clear user preference for number of competitors to rank, we decided to have Olympic audience members rank the *Top 5*. With the noise, excitement, and distractions at the Olympics, ranking the *Top 5* competitors may be easier and thus more fun for the audience.

Mechanisms for Audience Participation

We explored the use of different mechanisms for audience participation while considering usability issues, technical implementation details and cost factors.

Handheld devices, such as cell phones and Personal Digital Assistants (PDAs), were examined for use in our audience voting system. The small screen technology afforded by these devices could allow information to be delivered about the athletes and technical details about the sport and the judging criteria. We explored the feasibility of providing wireless capable PDAs for rent or loan to some or all audience members. It was determined that providing devices is prohibitive given the cost guidelines. As the artistic gymnastics and diving events overlap, separate sets of equipment are needed for both venues (15,000 seats and 6,500 seats respectively). At an estimated \$550 US per device, it would cost \$5.9 million to provide devices for only half the spectators.

Allowing users to vote with their own devices, although cost effective, presents serious user support and compatibility issues and more importantly would exclude many audience members who either do not own or did not bring a device. There are also several usability issues. Users with dexterity problems may find data entry via a stylus or small buttons difficult and the stylus is easily lost. Those with low vision may have to change their eyeglasses to view the competitors and then the small screen. Other usability factors associated with a diverse audience include literacy levels that may impact on understanding information and varying levels of comfort with technology. An international audience requires several languages and character sets. PDAs can be easily dropped (i.e. when standing for ovations) or have food/drink spilled on them. Batteries need to be recharged. Theft may be a problem.

Volunteers must be trained or technical support staff hired to provide user support for those with technical difficulties.

We also explored the use of a wireless response system (www.replysystems.com), which uses wireless numeric keypads. These devices cost less than handhelds as they are designed solely for input. The keypads contain finger-sized keys that require a single key press to submit a vote, eliminating the need for a stylus or awkward buttons. A strictly numeric keypad eliminates problems with language and internationalization. For those who are not technology-savvy, the keypads may be less intimidating. However, cost again proved to be the limiting factor. To support large venues, high-end (\$245 US) keypads are necessary. Even if only half the audience is provided a keypad the cost is \$2.5 million, excluding base stations and servers.

SOLUTION

With these issues in mind, we selected a low-tech method of audience interaction that will allow anyone to participate regardless of language or technical ability. Our solution consists of the *display of results* on large screens that provide a context for *visual audience interaction* through the use of ranking booklets and *capture of input* through analysis of photographs of the audience.

Maynes-Aminzade et al. [4] presented several design guidelines and principles in the areas of *system design*, *game design*, and *social factors*. *System design* guidelines suggest that the focus of the system be on the activity and not the underlying technology. In our system, we provide a low-tech input mechanism that allows audience members to easily submit their input without needing to interact with technology at all. The complexity lies in the capture and analysis of audience input. *Game design* guidelines suggest that audience activity should be varied with periods of non-activity. Our system has a strict timeline after each performance during which audience members must display their ranking. Once the voting window has passed, audience members are able to relax until the end of the next performance. *Social factors* guidelines advise “playing to the emotional sensibilities of the crowd” [5]. In our system, audience members have immediate feedback as they view their neighbours’ rankings. This will help create feelings of camaraderie and competition between audience members and will also instill trust in the process as the audience can clearly see the will of the majority in the sea of colour in the stadium. Tampering of the results would be difficult. Our approach is similar to Carpenter’s [1] who aimed to provide “a technique whereby one or a great number of people may simultaneously communicate information to a computer, using a simple device of minimal cost”.

Display of Results

We provide context for audience ranking through the display of previous results on large screens, standard in Olympic venues to display results and advertising. As

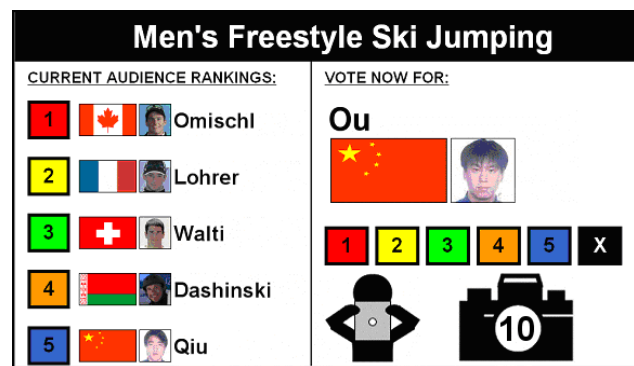


Figure 1. Mockup of large screen display

audience voting is part of the official event, we assume that our voting interface can be incorporated into the display schedule. A dedicated display is not required as voting takes only 15 seconds total.

The current *Top 5* rankings of the audience will be displayed (Figure 1), including the position and name of the competitor and the flag of their country. Although not shown on the mockup, all text will be in official IOC languages. The position will be displayed on a square, colour coded to match the ranking cards of the audience. This simple pattern matching scheme will enable a multicultural audience with varying levels of literacy to understand the effect of their choice. We also display an image of how to correctly hold the card to have the vote recorded. A countdown to the vote will be incorporated, counting down the ten seconds until the vote. The display will indicate when the vote has been recorded.

Visual Audience Interaction

Our approach requires that the audience rank the current competitor in relation to the previous competitors already ranked. There are six choices (placing the competitor within the *Top 5* or out of the running). For the current competitor to be considered in first place, the majority of the audience should rank that competitor as the new #1.

Our solution uses voting booklets (see Figure 2) that consist of six cards joined by rings forming a flipbook. Each card represents one of six possible votes for ranking the current competitor. Booklets are designed so that when an audience member is holding the appropriate colour of their ranking card outward, the ranking (1..5, and X) remains visible to the user. The estimated cost of producing each booklet is \$3-4 US. Voting booklets could become souvenirs for this Olympics. Images of Olympic mascots and sponsor logos could adorn the back of cards. A strap could be attached for easy carrying between events.

Capture of Input

Audience input will be captured with digital cameras suspended from the stadium roof, coordinated with the countdown on the large screen display. The image of the

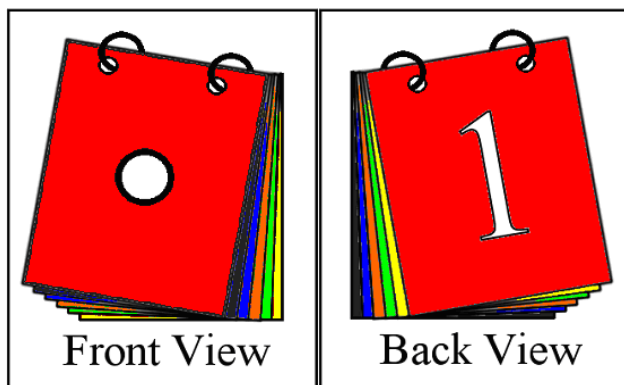


Figure 2. Mockup of voting booklet

vote will be analyzed to determine the audience's views. A colour histogram will be generated with bins for each possible ranking. This will be used to determine the median ranking of the competitor. The competitor will then be inserted into the appropriate position in the *Top 5* list or not included. The updated list can be displayed to the audience on the large screen for immediate feedback or, if the large screen is busy, at the next time to vote.

There are several technical details of the image analysis and vote recognition algorithm that must be finalized. Issues of lighting, audience movement, and background colour noise must be resolved to ensure that each vote is correctly captured. Baseline images of the stadium, its shadow patterns, and of the audience in a non-voting state will be used to filter the image before recognition. As it is necessary to not only recognize and distinguish between six colours on the ranking cards but to distinguish the cards from colours found in the audience, care must be taken when selecting the final card colours to ensure that they are unique and separated in the colour space used. A white dot will be centered on the card will aid in card identification. Sampling the surrounding pixels will give the colour of the card. If all the pixels are the same colour, we assume that a voting card has been found. The size of the cards and white dots will need to be calibrated when the distance from the audience to the camera(s) is confirmed.

Evaluation

We conducted a wizard-of-Oz evaluation to test the effectiveness of our solution. For the event, we used a video recording of a Men's Freestyle Aerial Jumping competition, an event with a format and scoring scheme similar to that for diving. Forty undergraduate students watched a round of jumps and after each jump, when prompted by the large screen display (see Figure 1), indicated their rankings using construction paper as cards. The actual judges' marks and placement for each skier were used for comparison with the audience results.

After the trial, participants completed a brief questionnaire of both positive and negative statements on a five-point

Likert scale (strongly disagree to strongly agree). Participants found the voting cards easy to use ($t(39) = 9$, $p < .001$), and the scoreboard effective in both coordinating voting ($t(39) = 4.58$, $p < .001$) and presenting results ($t(39) = 4.21$, $p < .001$). By contrast, there was no consensus on whether more time was needed to vote, or whether the voting procedure was a good fit for the sporting event. Our participants knew little about ski jumping on the whole ($t(39) = -7.37$, $p < .001$), leaving us unable to examine interactions between knowledge of the sport and attitude toward the prototype.

The images taken during this study will be used to evaluate and fine-tune our vote recognition algorithm to ensure that it captures the intent of the audience as accurately as possible. We plan to take repeated images of the audience during a large sporting event to help determine the variability of the background image. Field-testing in a large stadium will be necessary prior to deployment.

CONCLUSION

We have presented the design process for our solution for audience participation at Olympic Events that meets the competition requirements and results in a design that is visual and simple for the audience, removing the complexity from the interface and pushing it into the technology. Throughout the process, we considered design theory, consulted with experts and users, and conducted studies to ensure that our solution is usable.

While our solution proposes image analysis of cards displayed by the audience to determine the appropriate audience ranking, the audience responses could be collected in other ways (i.e. text messaging). Should budgetary constraints be lifted or technology costs lowered, our ranking mechanism could be used with other input devices.

ACKNOWLEDGMENTS

We thank Dr. Kori Inkpen and Dr. Carolyn Watters for their advice during the process, the WIFL and DVRG research groups for providing us a forum to discuss our ideas, and Dr. Norm Scrimger and Billy Biggs for technical guidance.

REFERENCES

1. Carpenter, L.C. (1994). Video Imaging Method and Apparatus for Audience Participation. US Patent # 5365266.
2. DeLeeuw, W.C., Knowlson, K.L. & Neeham, B.H. (2000). Method of Using Video Reflection in Providing Input Data to a Computer System. US Patent # 6088018.
3. Inselberg, E. (2003). Method and Apparatus for Interactive Audience Participation at a Live Spectator Event. US Patent # 6650903.
4. Maynes-Aminzade, D., Pausch, R. & Seitz, S. (2002). Techniques for Interactive Audience Participation. In *Proc. of Int. Conf. on Multimodal Interfaces*.