Multimodal Interfaces for Camera Phones

Stephen A. Brewster and Jody Johnston
Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
+44 (0)141 330 4966
stephen@dcs.gla.ac.uk www.dcs.gla.ac.uk/~stephen

ABSTRACT
Camera phones are now very common but there are some issues that affect their usability. These can occur because users look through the LCD to frame the image and can often miss the icons displayed around the edge that present important information about exposure, battery life, number of shots remaining, etc. This may lead to shots being missed or poorly exposed. We created a sonified luminance histogram to present exposure information, a sound cue to indicate memory card space remaining and a tactile cue for battery charge status. A user study showed that participants were able to use the sonified histogram to identify exposure successfully and could recognise the status of the battery and memory card well, suggesting that alternative forms of output could free-up the screen for framing the image.

Categories and Subject Descriptors
H5.2. [**User Interfaces**]: *Auditory (non-speech) feedback; Haptic I/O*.

General Terms
Design, Experimentation.

Keywords
Camera phone, sonification, tactile, luminance histogram.

1. INTRODUCTION
Camera phones are now very common and a camera is now a standard feature in most new phones. The quality of the cameras has begun to rival low-end dedicated digital cameras. In Japan, camera phone sales now exceed 50% of the market [3] and 12.5% of Japanese consumers who own a camera phone use it as their primary camera. Hayes [3] also predicts that by 2009 89% of mobile phones sold worldwide will include a camera. Features from digital cameras are migrating to camera phones, with auto focus, movies, optical zoom and flash becoming common. This suggests that other features will follow, making photo-taking on camera phones indistinguishable from standard digital cameras.

Many manufacturers focus on the resolution of the image the camera can capture, rather than the user-interface. There are some usability problems with camera phones (and standard digital cameras too) which can cause problems when taking pictures. The LCD is used to frame the image to be taken; the user looks through the LCD to see what he/she wants to photograph. Cameras present status and error information around the sides of the LCD, often using small icons. Figure 1 shows a screenshot of the Nokia N95 phone's camera application with icons showing information about the camera's state around the sides. Figure 2 shows a typical digital camera display with many icons on the screen representing different types of information, including battery life, number of shots left on memory card, exposure, ISO speed amongst others. Figure 2 also shows a luminance histogram (bottom right corner) that can be used to asses exposure. It can be hard to see all of this information when framing a photograph as attention is focused on the image to be taken and not on the icons. They may also be turned off as they can be distracting or hard to see due to the limited viewing angle of the LCD. The result is that photos can be taken which are badly exposed, blurred due to a low shutter speed or the memory card may be full so that the image cannot be saved.

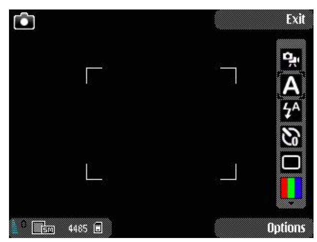

Figure 1: Nokia N95 camera application [6].

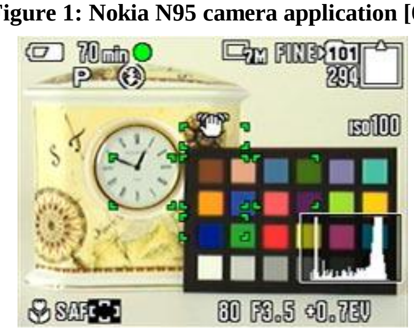

Figure 2: Display of a Sony DSC-V3 digital camera including a luminance histogram (from www.dpreview.com).

We conducted informal interviews with friends and colleagues who took pictures with digital cameras and camera phones to see if there were issues with camera interactions. A range of different problems were reported including: Blurred images due to too slow shutter speeds; Flash switched off by mistake; Running out of battery when not expected; Memory card filled up when in the middle of taking pictures; Wrong ISO speed set so images were too dark. We found other issues that we do not have space to dis-

MobileHCI 2008, September 2–5, 2008, Amsterdam, the Netherlands.

cuss here, but many of these were related to missed icons in the display of the camera; the camera does indicate many of these events using icons, but they were not seen by users as they were framing the picture they wanted to take. The aim of our research was to investigate if new forms of feedback could overcome some of the display issues and improve the usability of camera phones.

2. BACKGROUND

There is much work going on into the use of camera phones as interaction devices, for example using camera movements for gesture recognition [8], or identifying barcodes embedded in the environment to link to information [7]. However, there is little reported work going on into the design of the user interfaces to the cameras. In general, the interaction has been carried over from standard digital cameras. Audio is often used in those cameras, for example a beep to indicate focus lock and a shutter snap sound is commonly played to indicate that a photo has been taken. This is often the sound of the shutter on a film camera, played back as a sound sample.

Some very interesting work on photography and sound has been done by Frohlich [2] on *audiophotography*. His aim was to record audio alongside images taken to create a new form of photography. In this case the audio was from the scene being captured and not the interface of the camera.

Camera phones are unique as they have the capability to deliver audio and tactile feedback. Many phones have high quality audio hardware so that they can play music and games. This can be used to create sophisticated audio displays for the camera. They also include a tactile actuator for creating vibrations for alerts or incoming calls. Standard digital cameras do not have such features and do not allow additional software to be run on them, making the camera phone an interesting research platform for new interactions to improve the picture taking process.

3. MULTIMODAL FEEDBACK

Our interviews (and the display in Figure 2) showed that there were many different types of feedback that we could try to present multimodally. We chose three cues important for camera operation to see how they might benefit from alternative forms of feedback: the luminance histogram, memory card space remaining and battery charge status. The histogram is not yet common on mobile phones, but is likely to move over from standard cameras in the near future. Battery and memory card status are common and gave us a chance to investigate audio and tactile displays for range information of different types. These would also be applicable to other information displayed in Figure 2 (e.g. ISO speed or picture quality) so could provide some general support for camera design. We designed our feedback to avoid annoying the user or others nearby with excessive noise; E.g. where we used audio we replaced the existing camera sounds with our own new versions.

We developed our applications to run on a Nokia N93, the most sophisticated camera phone that Nokia produced at the time (February 2007). It had a 3.2 megapixel camera with autofocus and a 3 inch LCD screen [6].

3.1 Sonifying the Luminance Histogram

The luminance histogram shows the tonal range of an image (see Figure 3). On the x-axis the left hand side of the graph shows the black pixels (luminance of 0), the right hand side the white (luminance of 255). The y-axis shows the number of pixels with that value. The histogram is a key method for assessing the exposure of an image and is better than looking at the LCD screen as these often cannot clearly present the full tonal range of an image. A histogram can be seen in the bottom right of Figure 2. For more detail on histograms see www.sphoto.com/techinfo/histograms-/histograms.htm.

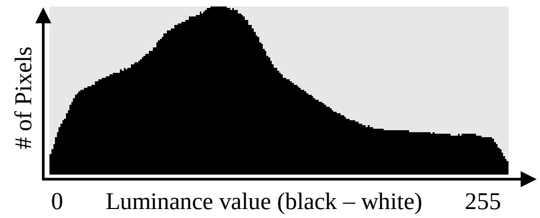

Figure 3: A luminance histogram shows the distribution of brightness within an image.

It can often be hard to see the histogram when framing an image and it takes up significant screen space, obscuring the image. As an alternative, some cameras can be set to flash the over exposed regions of a photo, i.e. those with a luminance value of 255 (though not commonly the underexposed parts), but this can be distracting to look at when composing the image and does not give all of the information that the luminance histogram provides.

To try to make the histogram more usable we sonified it. We took a simple approach, based on the Sound Graph [5], where we mapped the y-axis of the histogram to musical pitch and the x-axis to time. The histogram could then be played and a listener could hear the distribution of tones as a rising and falling pitch.

One limitation with this approach is that we would need to sonify 256 different values; one pitch for every brightness level in the image. Even if the sounds were played rapidly this would take a long time (if each sound was played for 100 msec. then the histogram would take 25 seconds to play). If the histogram took too long to play the image in the camera could have changed whilst it was playing, so the exposure information would be out of date. In addition it could become annoying to listen to such long sequences. We wanted to take particular care to ensure that any sounds used did not annoy the camera user (or others nearby).

A normally exposed image gives a histogram that is heavy in mid-tones and tapers down towards the edges. The most important values are at the far left and right; these are the ones that show the likelihood of the image being over or underexposed. There should be no peaks that get cut off or ‘clipped’ at the ends of the graph as that may mean information has been lost due to a poor exposure.

To reduce the number of sounds that we needed to play we focused the sonification on the ends of the histogram and binned the brightness values. We preserved more information at the ends of the histogram with less in the mid tones. After some informal experimentation we decided on 5 bins. This would mean that we could play the histogram rapidly but still give more information than just flashing the pixels with a value of 255. Our distribution of luminance values to the five bins was: 0-3, 4-26, 27-227, 228-251, 252-255. This gave the most resolution at each end of the graph, with the mid-tones compressed into one large bin.

Real-time MIDI synthesis was not possible on the N93 phone so we had to record samples for each of the sounds we needed and then play them back as appropriate. We restricted the pitches to a range of 8 notes spanning one octave (going up from middle C) to make the sonification sound as pleasant as possible without the need for very many sound samples. Each sound lasted 100 msec., this was short but still perceptible, meaning that the whole sonification lasted 500 msec. and could then be played on a half-press of the shutter button to give reliable exposure information.

To determine which note would be played for each histogram bin we first calculated the maximum pixel value of all the bins. Then for each successive bin, if the pixel value was less than $1/8^{th}$ of the maximum value it was assigned the lowest pitch of the 8 notes. If it was greater than $1/8^{th}$ but less than $2/8^{ths}$ it was assigned the second lowest pitch of the 8 notes, and so on. This created a set of 5 musical notes that varied in pitch to represent the histogram sequentially from left to right (low to high luminosity) as shown in Figure 4.

The sonified histogram was presented when the user half-pressed the shutter button, replacing the standard focus lock sound. He/she would then know if the picture would be exposed correctly before the picture was taken and could adjust exposure accordingly. It could also be used in a browsing mode where the user was skimming over the thumbnails of the images taken, where it can be hard to see if an image is exposed properly as it is very small, but a quick histogram would give some information about whether the image was worth keeping or deleting.

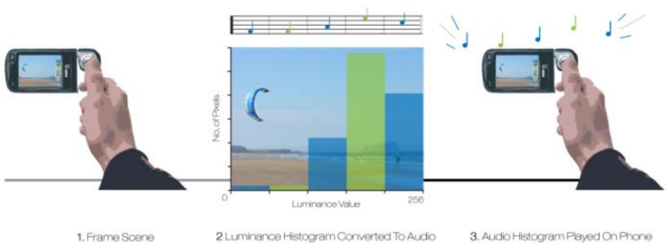

Figure 4: Sonifying the luminance histogram.

3.2 Sonifying Memory Card Space Remaining

Most mobile phones (and standard cameras) show the remaining number of photos that can be stored on the memory card. This is sometimes shown as a number or an icon (Figure 2 shows both a number and an icon top right). We investigated the presentation of this by altering the shutter snap sound. Some cameras already give an alert beep when the card is full; others just present a message on screen. By changing the snap sound to show the decreasing space available the user would get a clearer view of when the card will be full and no more pictures can be taken.

Figure 5: The cues used to indicate different amounts of free space on a memory card.

Our design used the standard shutter snap sound until there was space remaining for less than 25-30 more photos (the exact number varies due to the amount of compression of each image and the resolution used). At this point the standard shutter snap sound was low-pass filtered to create a more ‘muffled’ sound. This early warning sound was played whenever a photo was taken until there was space for less than 5-10 more photos. At that point the sound was changed again. This time the low-pass filtered sound was mixed with the sound of a slide whistle going down. This created the impression of something decreasing rapidly. Figure 5 shows the details. By changing the shutter snap sound we could give a simple warning of the state of the memory card in a way that did not change the behaviour of the camera at all. A finer level of detail could be presented to give a more precise measurement of remaining space, but it would be hard to make this perceptible and also it may not be necessary. The amount of space remaining only becomes really important when the card is nearly full.

3.3 Tactile Feedback for Battery Charge State

The final piece of information we delivered was the battery life remaining. This is commonly presented as a series of graphical bars on a mobile phone display. In Figure 2 the top left corner shows an icon of a battery that becomes more transparent as the charge is reduced. As we had used audio for the memory card cues we decided to use tactile feedback for this one as the actuator in the phone is little used for anything other than incoming calls or messages and could make a useful alternative display. Care must be taken when using the actuator that vibrations do not cause the camera to shake and cause blurred pictures. The audio display could have been used in this case, but we were interested in exploring the different display modalities available on the phone.

The idea for the feedback came from Brewster and King’s tactile progress indicator [1]. There, the time between two tactile pulses indicated the rate of a download, with a shorter time meaning less download remaining. We built on this with a longer time delay between pulses to indicate a fuller battery and a shorter one to indicate the battery was running down. The Nokia N93 phone has a 7 level battery indicator. We set the maximum time gap to be two seconds for a full battery, which meant that the final level (<=14% charge remaining) had a 0.25 sec. gap between the pulses. Each pulse lasted for 250 msec. and was played at the maximum output level of the actuator. The reduction in the time between the pulses has the effect of increasing the urgency of the cue (see Figure 6).

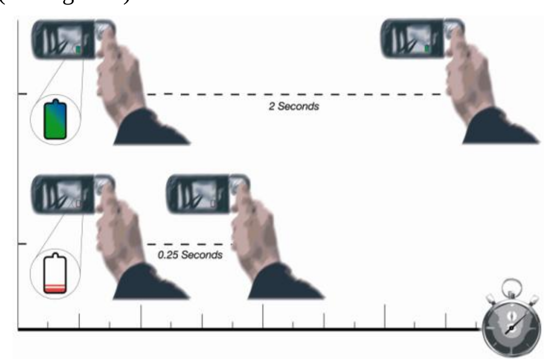

Figure 6: Tactile feedback to indicate the amount of battery charge remaining.

A key question was when to play this cue. It does not have a direct connection to picture taking like the memory card capacity; the phone will reduce in charge over time whether or not any photos are taken. We decided to play it directly after a photo was taken as our application in this study was photo taking, but it could be presented less frequently.

4. EVALUATION

We conducted an evaluation of the new display techniques to assess their effectiveness. We wanted to see if users could understand the exposure presented by the sonified luminance histogram as well as the visual one. We did this with a multiple choice within-subjects design, presenting one histogram and 3 images of different brightness. Each participant had to identify the image that the histogram represented. We created 30 images (10 underexposed, 10 overexposed and 10 normal) with each person performing both conditions (in a counterbalanced order) with the histogram presented either in a standard visual form (all 256 luminance values, as in Figure 3) or our sonification. At the start of the study we briefly trained people to use both the visual and audio histograms by giving them 4 sample photos (not used in the main test) that showed a range of different exposures and explained how both histograms worked.

For the memory card and battery status tests we gave the participants the phone and asked them to take a series of 20 pictures and after each gave either a tactile or audio cue. We gave a mixture of normal, medium and low alerts for both types of cue (for the battery cue we took full, medium and low battery. We could have tested all levels, but testing with just three reduced the time taken for the study and gave us enough information to know if users could get useful information from the vibrations) and participants had to record the value the cue represented on a response sheet. At the start of the study we showed the subjects the different ranges of tactile and audio feedback they would receive in the test.

We used 20 participants, all students from the University. Only four participants had used visual histograms before but all regularly took pictures with camera phones or digital cameras.

4.1 Results

The mean score for correct identification of the visual histograms was 24.3 out of 30 (SD 2.8) with 23.5 (SD 2.6) for the audio. This suggests that participants could match the exposure to the histograms well in both conditions, and that there was little difference between the two presentation methods. There were no differences between the identification rates of under-, over- or normally exposed images in either case. The results suggest that the five-bin histogram was effective at giving enough information to assess the exposure of an image correctly, without taking up the valuable screen space that a visual histogram uses.

Participants scored 86% correct identification for battery and 82% for memory card cues overall. The low battery/memory cues were particularly well recognised with 95% and 100% identification rates respectively, these being the key ones that users need to be aware of to keep taking pictures. This suggests that these cues could potentially replace their visual counterparts.

5. CONCLUSIONS

The sonified histogram was successful at presenting exposure information as users were able to recognise the exposure it presented with ~80% accuracy, close to the same level as the visual histogram. By presenting the histogram when the shutter was half-pressed they would be able to get a good idea of whether the picture would be well exposed or not and could correct it if desired by adjusting exposure compensation or re-framing the image. This would not annoy others nearby as the sounds would be little different from the normal focus lock sound.

We are investigating methods to sonify the full histogram so that full exposure information can be presented with a greater resolution. One solution could be to use High-Density Sonification [4], where all of the values in the histogram would be played in parallel. This method was successful for rapidly presenting rows and columns of data in large tables to blind people to give an overview of the data, so may be able to present the full histogram rapidly enough for our purposes.

The evaluation of the battery/memory feedback cues showed their potential for replacing the icons commonly overlaid on the viewfinder. By using the phone's standard vibration actuator for the battery cue and an altered shutter snap sound for the memory cue, the feedback could be presented in a way that is easily interpreted by a camera phone user.

Overall, the results show that some visual icons presented on the screen of a camera phone could be presented in other modalities. This would allow users to frame the images they want without distraction and still get feedback on important camera states, allowing them to avoid errors that could mean lost or poor pictures. It also shows that novel forms of feedback can potentially improve the picture taking process and that camera phones are a good platform to investigate novel interactions for photo taking.

6. ACKNOWLEDGMENTS

This work was funded by EPSRC Advanced Research Fellowship GR/S53244 and the Nokia University Donations Programme.

7. REFERENCES

[1] Brewster, S.A. and King, A.J., An Investigation into the Use of Tactons to Present Progress Information. In Proceedings of Interact 2005, (Rome, Italy, 2005), Springer, 6-17.

[2] Frohlich, D. Audiophotography. Kluwer Academic, Dordrecht, NL, 2004.

[3] Hayes, J. Mobile Imaging - Technology Trends, Consumer Behavior and Business Strategies. InfoTrends CAP Ventures, www.capv.com/home/Multiclient/ MobileImaging.html, Last accessed: 5/9/07.

[4] Kildal, J. and Brewster, S.A., EMA-Tactons: Vibrotactile External Memory Aids in an Auditory Display. In Proceedings of IFIP Interact 2007, (Rio de Janeiro, Brazil, 2007), Springer LNCS Vol 4663, 71-84.

[5] Mansur, D.L., Blattner, M. and Joy, K. Sound-Graphs: A numerical data analysis method for the blind. Journal of Medical Systems, 9. 163-174.

[6] Nokia, Nokia N93. Nokia, nseries.com/nseries/#l= products,n93, Last accessed: 27/2/2008.

[7] Toye, E., Sharp, R., Madhavapeddy, A., Scott, D., Upton, E. and Blackwell, A. Interacting with mobile services: an evaluation of camera-phones and visual tags. Personal and Ubiquitous Computing, 11 (2). 97-106.

[8] Wang, J., Zhai, S. and Canny, J., Camera phone based motion sensing: interaction techniques, applications and performance study. In Proceedings of UIST 2006, (Montreux, Switzerland, 2006), ACM Press, 101-110.