ASC: 03-1101; 1 May 2003

LESSONS LEARNED FROM CAESAR: A 3-D ANTHROPOMETRIC SURVEY

Kathleen M. Robinette and Hein Daanen*
Air Force Research Laboratory and
*The Netherlands Organization for Applied Scientific Research (TNO) Human Factors
kath.robinette@wpafb.af.mil

Abstract: The CAESAR project (Robinette et al. 2002) was the first anthropometric survey to provide 3-D human models. It was a multi-million dollar collaboration of more than 35 companies, several government agencies, and with representatives from 6 countries. Data were gathered in North America, The Netherlands, and Italy and two different 3-D scanning technologies were used. More than 13,000 3-D scans were provided and 4,431 subjects were measured. Given the complexity of such an undertaking you can imagine the problems that arose. This paper will discuss some of the most important issues, how they were resolved and any changes in how we would do this study if we were to do it over again.

INTRODUCTION

Planning for CAESAR began in 1992 with a workshop held in Dayton, Ohio (Vannier et al. 1992) and was followed by the establishment of a formal working group on 3-D Anthropometry with the mission of documenting the current technology and establishing the requirements of the survey (Robinette et al. 1997). A partnership for CAESAR was established in 1997 and the partner group met every 6 months to discuss progress, issues and application methods. Data collection itself did not start until 1998. The six years between the first workshop and the start of data collection was used to plan, get user feedback and test. This paper highlights some of the interesting lessons learned during the course of preparing for and conducting the survey.

INTERESTING THINGS THAT WORKED WELL

Many things went right because of good planning and early and continuous user feedback. Some of these were surprising. For example, consulting with the users of the data during planning helped to identify the need for seated scans. Originally, we had thought only about standing poses, but our data users noted that today most work is accomplished while a person is seated either in an office chair, a car, a truck, or some type of machinery. So this is the most important posture for many products. Identifying this early allowed us to have the 3-D scanners designed to accommodate the seated postures within the scanning volumes.

Another surprising result was the discovery that the users did not want the subjects to be measured nude, they wanted the subjects to wear their own undergarments. And the users most insistent were intimate apparel manufacturers. Their rationale was the nude body does not have the shape for which they design. Not only did this influence the choice of scanning garments, but we were also able to get the users to assist us in the design, testing, sizing, and manufacture of suitable scanning garments. A picture of the scanning garments on one of the female subjects is shown in figure 1. (The male subjects wore only the shorts.) These garments concealed the subject's own undergarments, while following the contours of the body with minimal tissue compression or bulk as a result.

Another part of our planning was testing every aspect, and this included many things other than just the accuracy and consistency of the scanner. It also included such things as testing garment color, subject pose, division of labor for the measuring team, and various methods for identifying key points called landmarks. One of the most important things that went right was the development of an effective method for identifying landmarks (Burnsides et al 2002). The method used stickers (and we also tested the color, size and shape of the stickers) to mark key landmark locations on the subject, special raised stickers for some points, and what we call a "semi-automated" method for finding and naming them in the scan. After spending many years trying to develop or find methods to automatically identify the location of important reference points on the body, it was determined that none of the methods would work consistently enough on all body types for use in a survey of thousands. Even using neural nets or other automated recognition packages to identify the location of the stickers seemed to only be successful about 70% of the time. This still required manual intervention for validation

Figure 1. Female subject in the scanning garments including the knit cap and landmark stickers.

and verification of the landmark location on all subjects. Instead, we planned to have a person view a 2-D (flat) image of the subject from the front and then from the back, and manually point to each sticker. The software then automatically located and named each point selected, did a heuristic check, and returned a file indicating any points that needed checking. The whole process took about 11 minutes per scan and had an accuracy of more than 98%, with the inaccuracies including stickers that were missing. This process was so effective it allowed us to identify and record real subject anomalies, such as a subject who had extreme scoliosis.

Another very important determination from pre-survey planning was the need for traditional measurements along with the scans. Several studies of scan extracted measurement consistency and comparability were conducted by people within the project as well as by other organizations. Software for automatically locating and taking tape measure type measurements such as circumferences and arcs that follow body contours was repeatedly found to be unreliable. (Daanen and Brunsman 1998, Paquette et al. 2000, Bradtmiller and Gross 1999). It was determined that the algorithms thus far developed for emulating a tape measure and the manner in which it would travel across a surface were not adequate. Only direct distance measurements taken between pre-marked points or between pre-marked points and rigid surfaces like the floor or the chair, were determined to be equivalent to, or better than, traditional measurement accuracy (Perkins et al. 2000). Since many of the traditional measurements have been used for many years, and since it may be many years before everyone has a 3-D scanner with the ability to identify pre-marked landmarks, it was felt it would be important to take some measurements the traditional way. As a result CAESAR included 40 measurements taken with calipers and tape measures and 59 point to point or point to surface measurements that were calculated from the scan points.

Another seemingly small thing now that it worked so well was the planning for things to go wrong. For example, we entered all data electronically and each subject carried his or her floppy disk of data, to and from each station. This was used to ensure that subjects' data did not get mixed-up when several subjects were going through the process at one time. In addition, for the demographics and traditional measurements we also filled out paper copies and these copies traveled with the subject on a clipboard as well. This turned out to be a lifesaver when we spilled coffee all over our laptop. We were able to continue data collection as if nothing had happened until we were able to replace the computer. It caused no delay in data collection at all. The paper copies also enabled us to cross check data that looked as if it may have been entered incorrectly later during data editing, therefore helping us ensure complete and accurate data were collected. We also brought along lots of spares of things such as data forms, floppy disks, and batteries and used common off the shelf spare parts or items whenever possible, which enabled us to run to a local store to purchase items that broke, or were lost. This minimized the number of delays in data collection that can happen when things go wrong.

INTERESTING THINGS WE WOULD DO DIFFERENTLY

Even with the best planning there will inevitably be things that go wrong during a survey. We were able to find a remedy for most problems, although some caused delays. One example was our recruitment of minorities. The fact that they are minorities means there are fewer of them in the population, and as our sampling strategy was to get an equal number of them we knew we had to work extra hard to find them. But we didn't realize we had to do some special things to get some of them to volunteer as well. We found that no matter how we tried to recruit people from some minorities we were not getting very many. With the help of the partners we contacted representatives of some of these minority groups for advice. We discovered that many American citizens who have been in the United States for many years do not speak English and therefore do not read English newspapers or watch English TV. So none of our advertisements or news stories were reaching them! When we were in Canada we had advertised in French and English, but in the US we also needed to advertise in Chinese, Korean, Vietnamese, and other languages and in the news media that people in these communities use. Since it was late in the survey we also offered an additional $50 incentive to volunteers so that we could get large numbers of the targeted minority groups quickly. We were able to get hundreds of minority volunteers within a couple of months. If we were to do this over again, we would do the multi-lingual advertising from the beginning and in addition the Asian languages and French we would advertise in Spanish as well, and we would talk more with different ethnic groups to find out the best way to recruit from their sector.

Several issues arose with the use of a different scanner. While we had thoroughly tested the Cyberware WB4 scanner for accuracy, for stitching the camera views, its ability to detect different colors and sizes of stickers, and the entire scan data collection process from start to finish, we didn't do the same testing on the Vitronic Vitus scanner purchased for The Netherlands portion of the study. The Vitus scanner had a very similar scanning methodology with lasers and digital cameras and top to bottom scanning etc. and even its appearance was very similar. Also, we had designed CAESAR so that it would not be scanner dependent. Unfortunately, we didn't know or had not been specific enough about how much alike the scanners had to be. It was the differences in the scanners that caused some difficulties we had to remedy.

First of all, when we had the Cyberware scanner built we specified that it come with software for viewing the data as a 3-D object within 5 minutes of a scan. (It actually is available within a few seconds.) We felt this was important for verifying that the scan was OK before the subject departed, and it proved to be very effective. There were several instances where something happened that we detected in this view and we were able to correct it in a re-scan within minutes. The Vitronic scanner did not come with this capability. It's software provided a flat 2-D picture of the view from each of the 16 cameras separately. As a result we lost some scans because of data problems that were not detected until weeks later when we were able to stitch the camera views together into one 3-D object. Furthermore, the software to stitch the images from the 16 cameras together did not come with the scanner, so we had to use a third party software for this and devise a scheme to ensure it was calibrated. (This was the first one they produced and they may have such software with it now.) We found that our scheme for calibrating the stitching of the images didn't seem to work consistently, months after the data were finished being collected we were still working on methods to stitch them effectively. We finally arrived at an acceptable method, but this delayed delivery of the 3-D models.

Secondly, the Vitronic scanner had a better resolution but a smaller scanning volume. In the seated position some body parts were outside the 3D-scanner pick up area and other parts were invisible for the color cameras , making it more difficult to detect the markers. Figure 2 shows a Vitronic scan and a Cyberware scan side by side. You can see that the Vitronic image has large white or overly light areas such as that in the middle of the subject's back that make it difficult to see the surface. This made identifying the white landmarks much more tedious and time consuming.

If we were to do this again we would put any new scanner through the same tests as we did the first scanner to determine the best lighting scheme, camera scheme, landmark detection scheme, scanning garment colors, sticker colors or shapes, whether to use luminance or color for landmark detection, image stitching accuracy and calibration etc. etc.

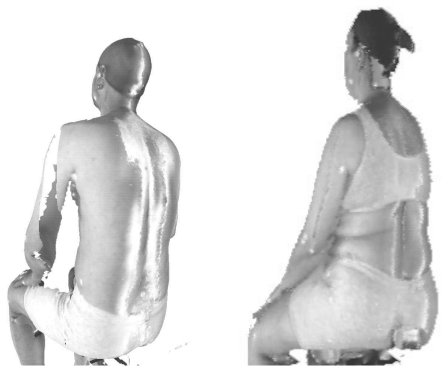

Figure 2. Vitronic scanner image (left) and Cyberware scanner image (right).

SUMMARY

In summary, some of the most important lessons for conducting a survey in our opinion are:

1) Have continuous user (customer) feedback and support from concept to delivery
2) Test everything before you start, including any new or different pieces of equipment no matter how similar they seem
3) Use multiple methods to recruit subjects, and consult with representatives from the different segments of your sampling strategy regarding recruitment
4) Be flexible and have a back-up plan for everything you can think of that might go wrong

REFERENCES

Bradtmiller, B. and Gross, M.E. 1999. 3-D Whole Body Scans : Measurement Extraction Software Validation. *Proceedings of the SAE International Digital Human Modeling for Design and Engineering International Conference and Exposition*, The Netherlands, PDF File 1892.

Burnsides, D.B., Boehmer, M. and Robinette, K.M. 2001. 3-D Landmark Detection and Identification in the CAESAR Project. in *Proceedings of the Third International Conference on 3-D Digital Imaging and Modeling Conference*, Quebec City, Canada, pp. 393-398, IEEE Computer Society, Los Alamitos, CA.

Daanen, H., and Brunsman, M.A 1998. *Difference between manual anthropometric measurements and anthropometric measures derived manually from whole body scans: a pilot study.* TNO-report TM-98-A004, TNO Human Factors Research Institute, Soesterberg, The Netherlands.

Paquette, S.J., Brantley, J.D., Corner, B.D., Li, P., and Oliver, T. 2000. *Automated Extraction of Anthropometric Data from 3D Images,* Apparel Research Network report (http://arn.iitri.org/docs/scan/systems/paquette.html)

Perkins, T., Burnsides, D. B., Robinette, K. M., & Naishadham, D 2000. Comparative Consistency of Univariate Measures from Traditional and 3-D Scan Anthropometry [CD-ROM]. *Proceedings of the SAE International Digital Human Modeling for Design and Engineering International Conference and Exposition*, Dearborn, Michigan, USA, PDF file 2145.

Robinette, K.M.; Vannier, M.W.; Rioux, M.; and Jones, P.R.M. 1997. *3-D Surface Anthropometry: Review of Technologies,* AGARD Advisory Report No. 329, Advisory Group for Aerospace Research and Development, 7 Rue Ancelle, 92200 Neuilly-Sur-Seine, France.

Robinette, K., Blackwell, S., Daanen, H., Boehmer, M., Fleming, S., S., Brill, T., Hoeferlin, D., and Burnsides, D., 2002. *Civilian American and European Surface Anthropometry Resource (CAESAR), Final Report, Volume I: Summary*, AFRL-HE-WP-TR-2002-0169, Air Force Research Laboratory, Human Effectiveness Directorate, Crew System Interface Division, 2255 H Street, Wright-Patterson AFB OH 45433-7022 and Society of Automotive Engineers International, 400 Commonwealth Drive, Warrendale PA, 15096

Vannier, M.W., Yates, R.E. and Whitestone, J.J. 1992. Proceedings of the Electronic Imaging of the Human Body Workshop, Crew Systems Ergonomics Information Analysis Center (CSERIAC), Dayton, OH.