APPENDIX B Avatar gesture library details

This appendix provides details about the format and creation of the avatar gesture library. It consists of the following three sections:

- Performance capture system
- Gesture sample file format
- Personal experiences

The last section relates our experiences recording the gesture samples. We include it because the effort involved in designing gesture variations and then capturing them may not be immediately evident.

Performance capture system

We built a performance capture system using the *Flock of Birds* position/orientation sensors built by Ascension Corporation [4]. The *Flock of Birds* uses one or more tethered magnetic sensors. We used between three and six sensors to record gestures.

The specific number of sensors required depends on the particular gesture being recorded and on the overall body movement desired in the recording. For our purposes, it was sufficient to record gestures involving only one arm. An arm gesture involving rotations of the shoulder, elbow and wrist joints can be recorded with four sensors. The sensors are attached to the torso, upper arm, lower arm and hand of the model. More sensors used around the body would provide more "lifelike" motion. By capturing the movement of the body in reaction to the arm gesture.

Data was recorded at a rate of 50 Hz. We used only the orientation data from the sensors since we use only joint angles to drive the avatar animation. This data is then translated into joint angles by finding the relative angles between adjacent segments along the limb.

After recording, we processed the data lightly. A local smoothing function was applied to the trajectories to remove noise in the data. Then the samples were aligned and trimmed manually. The duration of the gestures varied from 2 to 7 seconds.

File format

The performance capture data is processed lightly, so that will be suitable for animating an avatar, and stored in ASCII text files. The file consists of a header and then the animation data. A sample file is shown in Figure B-1. The header information includes the name of the gesture, the number of parameters with their names and values, the framerate for the animation, and the duration of the animation in milliseconds.

Following the header is the joint trajectory data separated into key frames. Key frames are separated by an asterisk. Each line provides the joint angle data for a single joint. The first token is the name of the joint, and it is followed by four floating-point numbers representing the quaternion rotation of the joint. The rotation is measured relative to the preceding articulation chain segment. That is, the angle is the relative rotation of a joint.

In describing the avatar animation, we follow the *h-anim* standard. H-anim specifies the names for segments and joints in an avatar model, the joint hierarchy and other conventions for the geometric and kinematic representation of the avatar. Only the joints that are involved in the movement are included in the file.

Experiences with gesture recording

Recording the animation samples for a gesture requires the following steps:

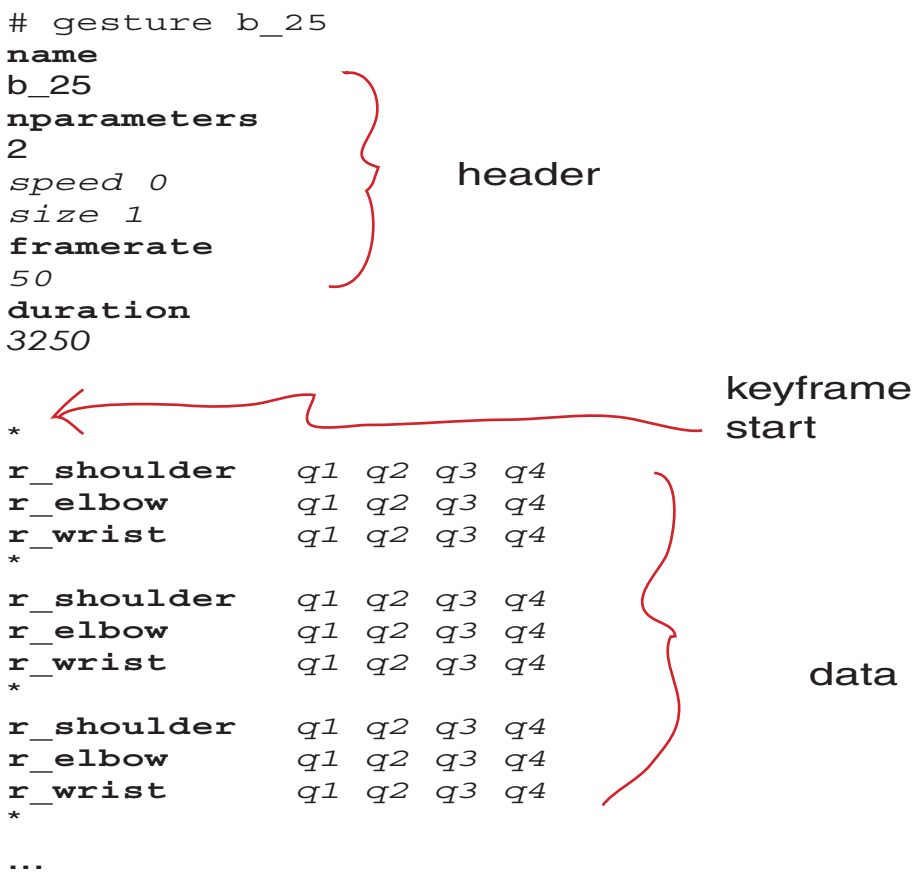

Figure B-1. Animation data file format

- Deciding on the basic (neutral expression) form of the gesture
- Designing the extremal forms for the gesture given the expression parametrization.
- Determining the appropriate placement of sensors on the gesture model, the person whose movement we recorded.
- Recording each of the samples.

We use one example of a recording session to describe our experiences with recording animations of communicative gesture. We will call our example gesture a *sweep* gesture. It is a single arm gesture. The arm initially starts at the side, is brought forward in front of the body, and

then is swept to the side away from the body. The neutral gesture is shown in Figure B-2. The

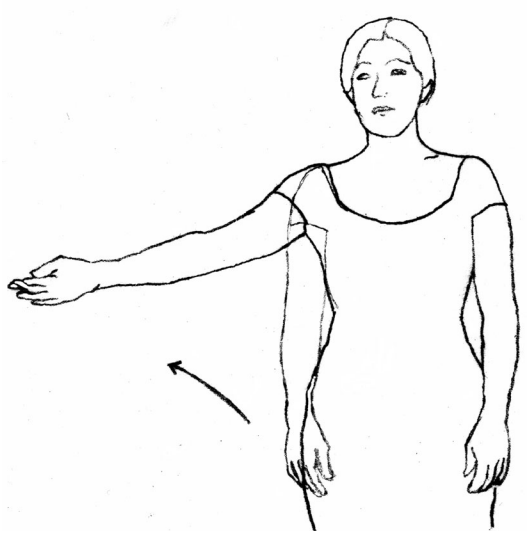

Figure B-2. Basic sweep gesture.

desired variation parameters are size and speed.

Deciding on the movements for the variations is a subjective judgement. Fortunately, we were using physical measurements as parameters. Even for our parameters there is a question of how small is a small version of this gesture, and how big is a big version. What is a neutral speed for a gesture? How fast is too fast? Figure B-3 shows the small and large versions of the sweep gesture that we decided on.

There were a number of complicating factors in designing the movement. First, all of the variations had to be similarly shaped, such as moving in the same direction, using the same joints, and having paths with the same number of inflection points. The last was particularly important because we use the animations as bases for interpolation. We needed the combinations slow-small, slow-big, fast-small and fast-big. Once we decided on what the variations

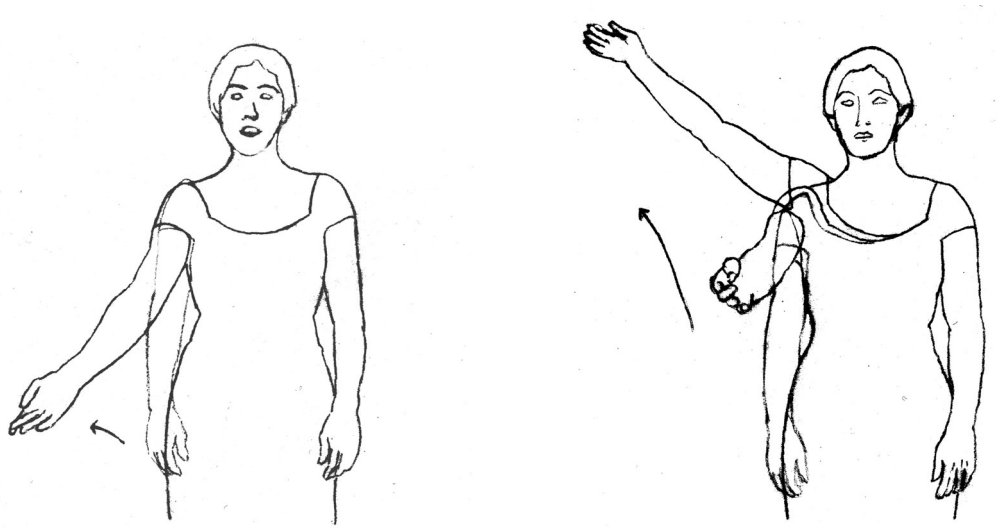

Figure B-3. A small version and a large version of sweep gesture.

should be, the gesture model had to decide how to perform them. They would make the gesture and together we would decided whether it felt right.

Finally, the recording introduced several other issues. For this example gesture, deciding where to put the sensors was simple. We knew it involved mainly the joints of one arm. Then we asked the gesture model to perform each of the movement variations. After each *take*, we looked at the animation on the avatar. Often we had to modify the movement because it looked quite different on the avatar which had much longer limbs than our model. Other times we had to do a retake the recording because the movement was not sufficiently similar to other recordings. At one point the sensors slipped, and we had difficulty placing them precisely in the same places they were before. This example shows there are a large number of environmental factors that inhibited the repeatability of the performance. Later we recorded the same gesture with three variations: size, speed and vertical placement.

Beyond this one experience, the modifications to movements required to express these variations will be different for different gesture types. For each gesture it must be decided if the variations are appropriate, and how they will be performed. Ideally, the gesture perfor-

mances that are recorded would come from observations of gestures that accompany real life conversations in a natural setting and would accurately record the range of expression available to that gesture. More likely, the person performing the captured gestures will be a gesture model, someone familiar with acting who can be involved in the design of the gesture variations.