The case for avatar makeup

Rosalee Wolfe
DePaul University
Chicago, USA
rwolfe@depaul.edu

Elena Jahn
Universität
Hamburg, Germany
elena.jahn@
uni-hamburg.de

Ronan Johnson
DePaul University
Chicago, USA
sjohn165@depaul.edu

John C. McDonald
DePaul University
Chicago, USA
jmcdonald@cs.depaul.edu

ABSTRACT
A challenge in creating a communicative sign language avatar is the visibility of the avatar's features. Compared to a face-to-face conversation with a physical signer, an avatar display occupies a smaller portion of a recipient's field of view, making the fine features and subtleties of sign language difficult to perceive. This paper discusses recent efforts to make facial nonmanual signals more visible to a viewer.

Author Keywords
Sign Language Avatars; Texture mapping; Sign Synthesis.

ACM Classification Keywords
H.5.m. Information interfaces and presentation: Miscellaneous;

MOTIVATION

One of the challenges in creating a communicative sign language avatar is the visibility of the avatar's features. Compared to a face-to-face conversation with a physical signer, an avatar display occupies a smaller portion of a recipient's field of view, and so appears smaller, making the fine features and subtleties of sign language difficult to see. The problem is exacerbated when the avatar is not displayed on a large desktop computer screen, but on a small mobile device. When such a device is held at the usual viewing distance, an avatar displayed on the device will appear smaller than a physical human viewed at a distance of 35 feet.

Subtle and not-so-subtle sign language processes can be difficult to see at long distances. Members of the Deaf community are well aware of this fact, and at an event, they will often arrange their seating to be near the interpreters so as to better see the signing. Facial nonmanual signals are particularly vulnerable to this problem, since the features of the face are smaller in size than the arms or the hands, as is demonstrated in Figure 1. The image on the right is a picture of an avatar as seen from a distance. Although the handshape might still be discernable with effort, there is little information left for the viewer to perceive on the face.

This poses a challenge for conventional computer graphics. Using traditional modeling techniques to create the geometry along with computationally efficient local lighting model is not adequate. They do not sufficiently support the transmission of finer grained facial changes. The shading effects achieved through either modeling fine geometric details via displacement mapping or modifying the surface normal via bump mapping to simulate such detail are too subtle to be seen at a distance. These details often contain important linguistic information that a viewer must be able to see in order to understand the utterance being produced by an avatar.

Figure 1: the vulnerability of facial nonmanual signals to obscuration by distance

A SOLUTION "OUTSIDE OUR BOX"

For inspiration, we look to traditional theater performance since actors have long had to grapple with the challenge of being seen by spectators sitting dozens of meters from the stage. Spectators may not be able to recognize or understand the actor's expressions due to the blurring of facial features. To counter this, actors use makeup not only to communicate their character but also to ensure that the entire audience will be able to recognize the actions of their character. When seen from a short distance, theatrical make up is exaggerated to a point where actors look unnatural. But from a distance, it emphasizes exactly those lines the audience needs to grasp the changes occurring on the face. Typically, eyes are surrounded by very bold dark lines, which enables the audience to see the positions and movements of the eyes from a distance. Theatrical makeup also enhances the cheekbone, which is one of the most movable regions of the face. This allows the audience to see the changes in the actor's expression.

Furthermore, makeup for an actor is a necessity because modern stage-lighting systems may remove color from a performer's face and will fill in and thus eliminate shadows and lines. Makeup restores the color and defines the facial contours to communicate an appearance appropriate to the characters the actors play. Makeup that makes a character look "friendlier" can add to its appeal and acceptability.

Although prosthetics or other SFX makeup techniques can create three-dimensional changes to the appearance of a stage actor, more often than not actors and makeup artists rely on the principles of light and shade to create the illusion of three dimensions [4]. The tools found in a theatrical makeup kit includes makeup base colors, rouges, colored liners for shadow and highlighting effects, eye makeup and false eyelashes. Creating the effects of stage makeup requires a digital equivalent for these tools.

THE CHALLENGE

Clearly portraying the face of an avatar requires consideration of multiple concerns, including

- emphasizing facial features
- giving prominence to the contouring of the face
- drawing attention to the folds of the skin
- creating wrinkles that are visible from a distance.

Not only are these effects necessary, but they may need to occur simultaneously. Creating these effects concurrently requires a new representation of texture maps for their implementation. This is analogous to the reliance of makeup artists on coloration applied to an actor's face rather than adding three-dimensional changes using SFX makeup techniques. Furthermore, representing these for avatar display requires developing techniques to control the timing of the appearance of these effects. The remainder of this presentation will address a new representation and suggests methods for their control.

PRIOR EFFORTS

Alternatives for modeling fine facial details include displacement mapping [3], bump mapping [1] and texture mapping[2]. Using displacement mapping on an avatar to change the 3D geometry is analogous to a makeup artist's use of prosthetic makeup to change the appearance of an actor's face. Unfortunately, for an avatar, this technique increases computing requirements considerably, and the results are not visible from a distance. A method that requires less compute time is bump mapping, which manipulates the surface normal. When the lighting calculation uses the altered normal, the visual effect is a perceived perturbation of the surface, without the expense of adding the many small polygons required by displacement mapping. It doesn't actually change the geometry, but it dynamically adjusts the interaction with the light as if geometry deformations were there. Unfortunately, when used to create creases or folds in a face, this approach also produces effects that are too subtle to be perceived under usual viewing conditions.

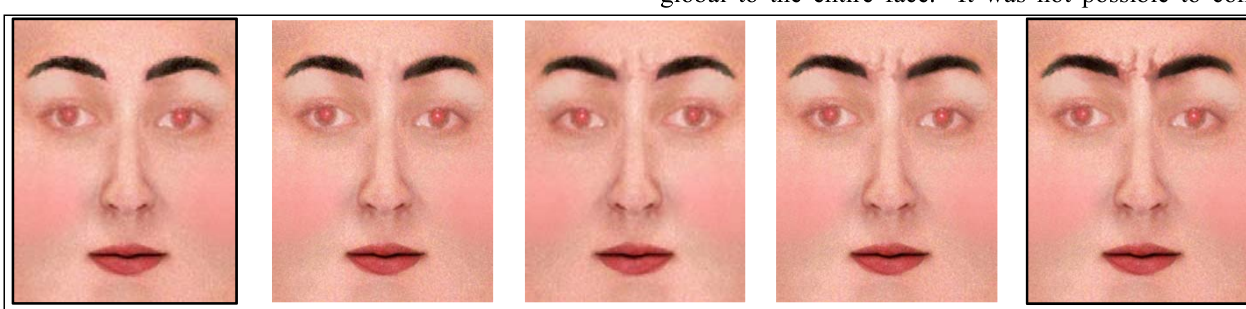
In contrast, texture mapping has the potential to create visual effects that are visible at a distance. In texture mapping, a two-dimensional image is applied to the 3D geometry as a part of the rendering process and thus has a direct analog to the base colors, rouges and liners used by makeup artists in the theatre. In other words, texture mapping "paints" a 2D image onto the geometry.

Texture mapping has an advantage over theatre makeup – it can change appearance over time during animation of sign language utterances. By structuring the final desired texture not as a single image but as a set of images, each with a transparency control, it is possible to cause visual effects to appear and disappear on the avatar. This is similar to the layered structure used by Adobe Photoshop [6] to organize imagery. As Figure 2 demonstrates one can mix the two layers to create an effect that can be dynamically changed in time. As the opacity of the second image increases, it becomes more prominent in the texture.

Schnepp [5] used this technique to create forehead furrows when the avatar lifted her eyebrows to pose a Yes/No question in ASL. He controlled the appearance of the furrow by calculating an opacity factor for the foreground simply based on the depth of the eyebrow lowering:

```
p = (browZ – browZmin)/(browZneutral – browZmin)
p = clamp (p, 0, 1)
image = fg * p + bg * (1 – p)
```

There was a limitation to this approach. The effect was global to the entire face. It was not possible to combine multiple effects. For example, if the avatar were to pucker her lips while asking a WH-question, the previous approach



Background image 0% brow furrow	30% brow furrow 70% background	50% brow furrow 50% background	70% brow furrow 30% background	Foreground image 100% brow furrow

Figure 2: Imagery organized by layers

would force an animator to choose between the pucker or the brow furrow.

A NEW APPROACH

This presentation proposes a different organization of the texture maps to make it possible to combine multiple sets of

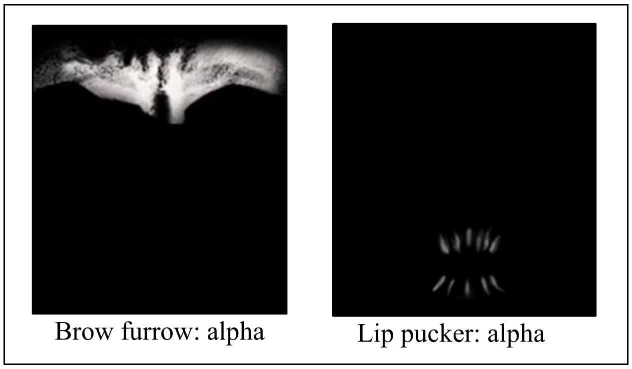


Figure 3: The alpha channel for brow furrow, lip pucker

visual effects simultaneously rather than being forced to choose just one. This new organization has a regularized form and can easily computed on today's GPUs. With this new format, textures can also be used in place of fine detail while adding prominence to features such as the eyes. All of these can occur concurrently.

To the conventional red, green and blue channels of an individual texture map, this approach exploits the alpha channel. The alpha channel controls the visibility of individual pixels within the texture map. A value of zero for the alpha channel of a pixel renders the pixel invisible. For "wrinkle" layers, most of the texture map will be invisible except for areas there the wrinkles occur as demonstrated in Figure 3.

For each layer, in addition to the global visibility value as in [6], the local visibility information in the alpha channel allows multiple layers to affect the appearance of the texture map. Figure 4 demonstrates the effect of multiple layers, and the result, as rendered on the avatar is shown in Figure 5.

RESULTS AND NEXT STEPS

This layered approach to texture mapping has been implemented to as a shader in GLSL and runs at video rates on mid-range graphics hardware. Future plans include incorporating device responsiveness and resolution awareness in the approach, to adjust the intensity of the texture map based on the resolution of the display.

Figure 5: Final rendered result

REFERENCES

[1] Blinn, James. Simulation of wrinkled surfaces. *Computer Graphics*, 12, 3 (July 1978), 286-292.

[2] Blinn, James and Newell, Martin. Texture and reflection in computer generated images. *Communications of the ACM (CACM)*, 19, 10 (October 1976), 542-547.

[3] Cook, Robert L. Shade Trees. *Computer Graphics*, 18, 3 (July 1984), 223–231.

[4] Corson, Richard. *Stage Makeup*. Talylor and Francis, Milton Park, 2018.

[5] McDonald, John, Wolfe, Rosalee, Schnepp, Jerry et al. An Automated Technique for Real-Time Production of Lifelike Animations of American Sign Language. *Journal of Universal Access in the Information Society*, 15, 4 (2015), 551-566.

[6] Oh, Byong Mok, Chen, Max, Dorsey, Julie, and Durand, Frédo Durand. Image-based modeling and photo editing. *Computer Graphics*, 35, 3 (August 2001), 433-442.

Background Brow furrow Lip pucker Added together

Figure 4: New organization combines multiple layers of detail.