Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology (SLTAT 7), pages 67–72
Language Resources and Evaluation Conference (LREC 2022), Marseille, 20-25 June 2022

A Novel Approach to Managing Lower Face Complexity in Signing Avatars

John C. McDonald, Rosalee Wolfe, Ronan Johnson
DePaul University
Chicago, USA
jmcdonald@cs.depaul.edu, {rwolfe,sjohn165}@depaul.edu

Abstract

An avatar that produces legible, easy-to-understand signing is one of the essential components to an effective automatic signed/spoken translation system. Facial nonmanual signals are essential to natural signing, but unfortunately signing avatars still do not produce acceptable facial expressions, particularly on the lower face. This paper reports on an innovative method to create more realistic lip postures. The approach manages the complexity of creating lip postures, thus making fewer demands on the artists making them. The method will be integral to our efforts to develop libraries containing lip postures to support the generation of facial expressions for several sign languages.

Keywords: signing avatars, sign language representation, computer animation

1. Introduction

To improve deaf accessibility, multiple efforts have explored automatic translation from spoken to signed language. However, since signed languages have no widely accepted written form, any output from machine translation will necessarily require a display on a computer-generated human form. One of the most promising methods is a signing avatar, and while efforts to utilize avatars have been ongoing since the late 90's, the acceptability of signing avatars in the Deaf community has been lukewarm at best (Austrian Association of Applied Linguistics, 2019).

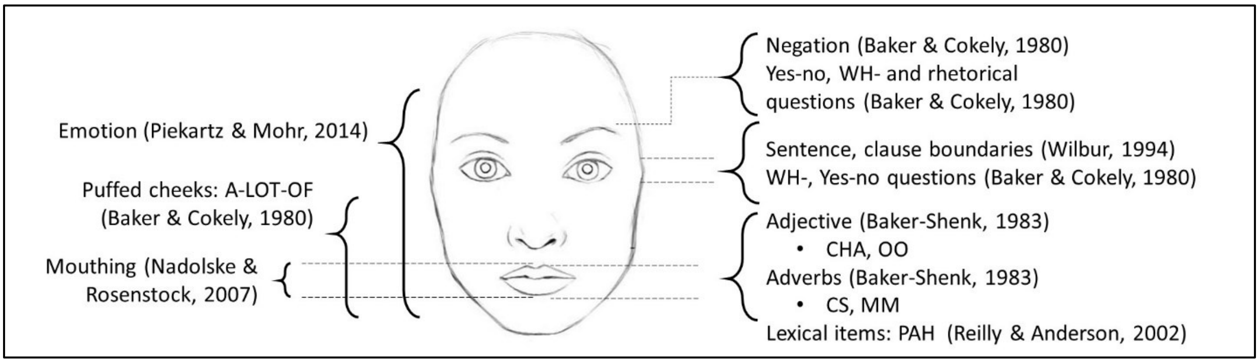

Emotion (Piekartz & Mohr, 2014)
Puffed cheeks: A-LOT-OF (Baker & Cokely, 1980)
Mouthing (Nadolske & Rosenstock, 2007)
Negation (Baker & Cokely, 1980)
Yes-no, WH- and rhetorical questions (Baker & Cokely, 1980)
Sentence, clause boundaries (Wilbur, 1994)
WH-, Yes-no questions (Baker & Cokely, 1980)
Adjective (Baker-Shenk, 1983)
• CHA, OO
Adverbs (Baker-Shenk, 1983)
• CS, MM
Lexical items: PAH (Reilly & Anderson, 2002)

Figure 1: No linguistic or extralinguistic process has an exclusive franchise on a facial feature.

One of the primary criticisms from the Deaf community has been the lack of adequate motion on the face, including the lack of adequate mouthing (Verlinden, Tijsseling, & Frowein, 2001), (Kipp, Nguyen, Heloir, & Matthes, 2011), (Ebling, et al., 2015). This paper revisits the existing technologies for mouthing on human avatars and proposes a novel approach that is tuned to the unique requirements of sign language, allowing for greater expressivity, and imposing fewer demands on the artists creating signed discourse.

2. Background

The technology of applying mouthing to a signing avatar draws on the traditional lip sync process used in character animation (Parent, King, Fujimura, & Osamu, 2002). The process involves four steps:

1. Generate phonemes corresponding to a spoken word.
2. Map each phoneme to a viseme, which is the phoneme's visual appearance.
3. Retrieve facial poses (or settings) corresponding to each viseme from a library.
4. Apply facial poses to the avatar as animation keys.

A prerequisite to this process is the creation of a library of visemes. Creating a realistic set of facial postures to portray visemes is a difficult and time-consuming task that does not always yield satisfactory results (Brumm, Johnson, Hanke, Grigat, & Wolfe, 2019). This paper describes an innovative approach to viseme creation that manages the complexity of the process in an animator-friendly way. The approach is sufficiently general that it also supports the creation of postures for mouth gestures as well as for visemes.

There are two main approaches to creating visemes: using morph targets[1] and using muscle simulation. Morph targets have the advantage of simplicity (Alexa, 2002). To create a library of visemes, artists manually sculpt each viseme from

[1] Another term commonly used in the animation industry is "blend shapes".

a copy of the original model and can utilize their favorite sculpting tools. From a software development standpoint, morphing is straightforward to implement. However, the same simple implementation can create unanticipated effects. All changes in position in morphing follow a linear path, which is not compatible with human facial anatomy. Additionally, there is a deeper concern because in sign language, no one linguistic or extralinguistic process has an exclusive franchise over a facial feature and multiple processes can co-occur. With a morph implementation, multiple morphs will directly affect the same regions of the face simultaneously, but in an additive manner. The resulting effects are not natural-looking. Finally, from an implementation standpoint, morph representations require extensive in-memory storage. This is not necessarily a problem in desktop environments, but it can become a consideration on mobile devices.

3. Previous Work

Previous signing avatars have used both the morph-based (Jennings, Elliott, & Kennaway, 2010), (Kipp, Heloir, & Nguyen, 2011) and the muscle-based (Wolfe, et al., 2018) approaches, but feedback from deaf communities indicated that the mouth postures were not satisfactory. These avatars relied on the MPEG-4 H-Anim standard for manipulating the mouth (Ostermann, 2002). In the standard, there are 28 landmarks available to control lip postures. This was sufficient for early interactive agents to demonstrate the approach, but a rig capable of accurately portraying lip postures required more landmarks. Johnson (2018). developed a rig with 44 landmarks instead of 28. This facilitated smoother lip postures and made it possible to portray a wider variety of mouth postures than with the original H-Anim landmarks.

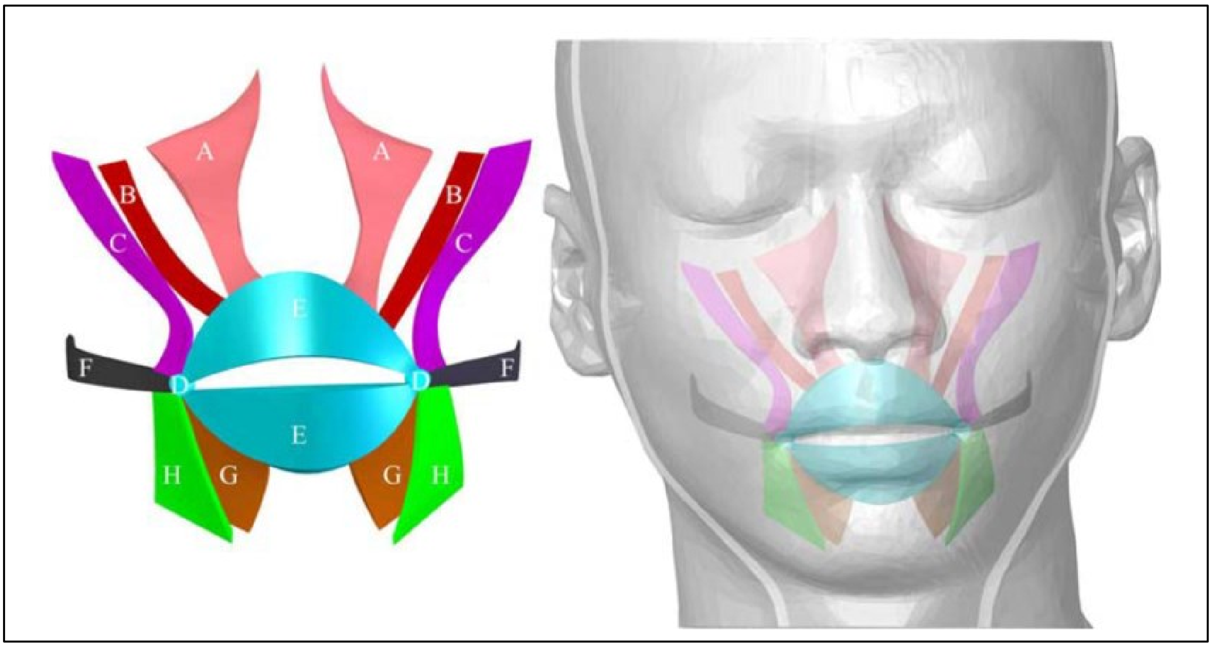


Figure 2: Selected muscles affecting lip shape, including levator labii superioris (A), zygomaticus minor (B), zygomaticus major (C), orbicularis oris (E), risorius (F), depressor labii inferioris (G) and depressor anguli oris (H) (Chen, et al., 2012).

The alternative to morph-based systems are muscle-based systems. Park and Waters (2008) examined facial structure beneath the skin and developed a parametric representation to simulate muscles. A muscle-based approach has the advantage of producing more natural results, as the underlying representation more closely mimics the muscle behavior in a human face. A distinct disadvantage of this approach is the increased burden placed on an artist using the system. A case in point is simulating the orbicularis oris to control lip shape.

The orbicularis oris is a complex multi-layered set of muscles that attach to the upper and lower lip. Researchers point out that, although anatomically it is a single muscle, from a functional viewpoint it actually consists of several components that either act independently or in concert with other facial muscles (Jain & Rathee, 2021). Figure 2 displays a simplified schematic of 10 of the 20 muscles attached to the orbicularis oris.

4. An Improved Approach

With the new capabilities for precision and a wider range of expressive possibilities came problems with usability. From an animator's perspective, the new rig was a step backwards. Instead of working with 28 landmarks to manipulate the face, the animator was confronted with the prospect of 44 landmarks to manipulate. In this state, the new workflow made it more difficult, not less difficult, to create believable mouth poses.

To counter this problem, Johnson began by organizing the facial muscles into groups, based on the perceived effect each group has on the face. He characterized the effect of various muscle groups on the lips, with the goal of making the lip posing process more manageable. Not surprisingly, the orbicularis oris is a member of each group. The other muscles in a group create localized changes to the geometry of the orbicularis oris. For a discussion of building the muscle representation, please see (Johnson, 2022).

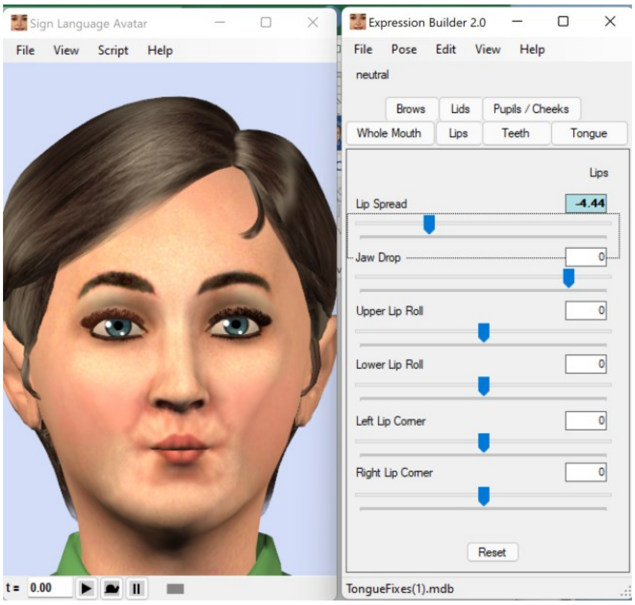
The muscle groups are attached to controls in the user interface in DePaul's Expression Builder (Schnepp, Wolfe, McDonald, & Toro, 2013). Each control is simply a slider, and there is one slider for each muscle group. The first six groups listed in Table 1 appear in the Lips panel, as seen in Figure 3. (The second six groups appear in the Teeth panel of the interface.) Per Table 1 all the sliders involve the orbicularis oris. Most of the sliders also manipulate connecting muscles that in turn affect the orbicular oris. In all, an artist has access to twelve sliders to manipulate the lips. This compares quite favorably to the 28 H-Anim landmarks and certainly a better approach than requiring the manipulation of a set of 44 landmarks. Artists can use this system to create not only visemes suitable for mouthing, but also postures for mouth gestures.

	Effect	Cooperating muscle group	Layer
1	Lip Spread	left/right risorius, left/right buccinator, obicularis oris	1
2	Jaw Drop	left/right depressor labii inferioris, mentalis, orbicularis oris	2
3	Upper Lip Roll	obicularis oris	4
4	Lower Lip Roll	left/right mentalis, left/right depressor labii inferioris, obicularis oris	4
5	Left Lip Corner	left Zygomaticus major, left Depressor anguli oris, obicularis oris	3
6	Right Lip Corner	right Zygomaticus major, right Depressor anguli oris, obicularis oris	3
7	Show Upper Teeth	left/right zygomaticus minor, left/right levator labii superioris alaeque nasi, obicularis oris	5
8	Show Lower Teeth	left/right depressor labii inferioris, left/right mentalis, obicularis oris	5
9	Left Upper Snarl	left levator anguli oris, left levator labii superioris alaeque nasi, obicularis oris	6
10	Right Upper Snarl	right levator anguli oris, right levator labii superioris alaeque nasi, obicularis oris	6
11	Left Lower Snarl	left depressor labii inferioris, left depressor anguli oris, obicularis oris	6
12	Right Lower Snarl	right depressor labii inferioris, right depressor anguli oris, obicularis oris	6

Table 1: Muscle groups and their effect on the lips.

5. Implementing the approach

The last part of this approach involves developing the infrastructure required to manage the behavior of the 44 landmarks in response to the manipulation of the sliders. Consider a single slider "Lip Spread" from the interface. It controls the effects of the buccinator muscles, which compress the orbicularis oris and the risorius muscles which spread it. Moving the Lip Spread slider to the left activates the buccinator which puckers the lips. Moving the Lip Spread slider to the right activates the risorius which widens the mouth.



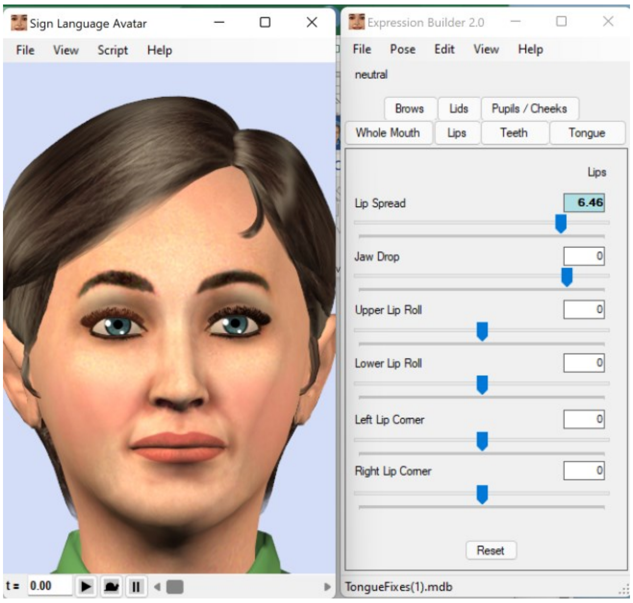
neutral

Lip Spread -4.44
Jaw Drop 0
Upper Lip Roll 0
Lower Lip Roll 0
Left Lip Corner 0
Right Lip Corner 0

t= 0.00

TongueFixes(1).mdb

A negative Lip Spread compresses the lips.



neutral

Lip Spread 6.46
Jaw Drop 0
Upper Lip Roll 0
Lower Lip Roll 0
Left Lip Corner 0
Right Lip Corner 0

t= 0.00

TongueFixes(1).mdb

A positive Lip Spread widens the lips.

Figure 3: Artists use sliders in the interface to create facial postures

5.1 Basic algorithm

Now consider a single landmark of the 44 landmarks on the mouth. The landmark has three positions of interest – one when the lips are fully pursed, one when the lips are fully spread and one when the lips are in the neutral position.

These three points define a path. Instead of following a straight line from one extreme to neutral to the other extreme, as in a morphing approach, the transition path follows an arc which approximates the local contour of the head. This creates a natural-looking transition, with no awkward or unnatural intermediate positions.

A script within the landmark connects[2] the landmark's position to the user interface slider "Lip Spread". The slider value controls the landmark's position along its path. Each of the 44 landmarks has a customized script connected to the "Lip Spread" slider that controls its path movement. When an animator adjusts the slider, the landmarks all move in concert, see Figure 3. The same scalar from the slider controls the rotations for all the landmarks.

Thus, the slider value is the parameter for all of the landmark scripts. This requires creating a strict consistency in the slider values in the user interface. For each parameter, a value of zero corresponds to the neutral position on the landmark. For symmetric sliders, the range is always -10 to 10. For asymmetric sliders, the range is always 0 to 10. (The jaw drop slider is slightly different for historic reasons, but its neutral position is at zero.)

Adhering to this consistency results in shorter scripts, and quicker code development. Further, resetting the entire face to neutral is simply a matter of setting all of the values of the user interface sliders to zero.

We used the scripting capability of a commercial animation package to prototype the approach. Figure 4 gives the pseudocode script for a landmark controlled by the Lip Spread slider. The initial statement designates the user interface slider as the connecting parameter; the second line ensures that the incoming parameter absolute value is no more than 1. The if statement distinguishes between the spread (positive) and pucker (negative) cases. The slerp (spherical linear interpolation) function calculates the angle of rotation for the landmark. Note that a single slerp operation between qmaxPucker and qmaxSpread cannot in general be assumed here because the half-way point between the two rotations may not be identity.

Some landmarks, particularly those on or near the center line, will be influenced by multiple muscles, and each muscle will have a different behavior. Scripts can accommodate this situation and blend the influences to derive a smoothly changing transition.

```
dependsOn LipSpreadSlider
t = LipSpreadSlider / 10
if t >= 0 then
 slerp identity qmaxSpread  t
else
 slerp identity qmaxPucker -t
```

Figure 4: Pseudocode to control landmark position from user interface.

5.2 Organizing multiple influences

As is demonstrated in Figure 4, the scripts for controlling a single cooperating muscle group are straightforward. However, on the human face, the orbicularis oris has multiple influences from many sets of cooperating muscles. Attempting to incorporate all influences into the landmark scripts would become unmanageable.

To accommodate the many influences while keeping script complexity under control resulted in a layered organization. Instead of having a single set of 44 landmarks, there are six sets of landmarks, one each for spread, jaw drop, lip roll, (mouth) corners, show teeth and snarl. Table 1 lists their layer assignments, with smaller layer numbers being more global (proximal) in the hierarchy. Each layer has its own set of scripts, and the complete lip posture is a result of multiplying the transform matrices of the corresponding landmarks in each layer.

5.3 From prototype to production

Our avatar modeling, rigging, and texturing occur in several commercially available animation packages (3ds Max, Maya, Substance Painter), and a custom exporter package converts these into a format compatible with our real-time avatar display. Likewise, the scripts connecting the user interface to the landmarks originated in a commercial package and needed to be exported to the real-time system. This presented a knotty problem, because the complexity of the scripts would require the addition of a parser to export them.

As an alternative, we added a specially formatted comment line at the beginning of each script. A colon-delimited line specifies

- Number of muscles influencing the movement
- Names of slider controlling the movement
- Whether the range of the slider control is symmetric or asymmetric
- Extreme maximum value of the landmark rotation (positive values of slider)
- Extreme minimum value of the landmark rotation (negative values of slider)
- Normative factors to convert incoming parameters from sliders to range from -1 to 1 or 0 to 1, depending on whether the parameter range is symmetric or asymmetric
- Weights corresponding to the influence of each muscle

For the pseudocode in Figure 4, the pseudo comment line would be

```
--:1:LipSpreadSlider:symmetric:qmaxSpread:
qmaxPucker:10:-10:100:100
```

Please note that the strictures of the paper format required a line break. Thus, the exporter only had to consider the first (comment) line of a script when exporting it. Given the adherence to a consistent writing style for the scripts, we were able to express the intent of the scripts in the form of a comment line. The exporter required only a few additional

[2] Commercial animation packages refer to these as "wires".

lines of code to process the scripts, and a single, generalized shader in the real-time avatar display accommodated all the scripted behavior.

6. Controlling motion

Posture is, of course, only part of the equation when dealing with avatars, since much of what distinguishes natural vs. robotic signing is carried in the motion between the animation keys. Thus, intuitive control of the interpolation between animation keys is critical, and the new facial bone structure and control set has several distinct advantages over both the MPEG-4 H-Anim localized control and the original Paula rig:

- A more compact control count (56 as opposed to 800) affords more space in the database for velocity/acceleration control,
- Each of the controls is a single scalar rather than a 3D position or set of Euler Angle rotations, so a single animation control can affect all of position/rotation information encoded in each script. Interpolating a scalar is more straightforward than interpolating a position and certainly more straightforward than interpolating a rotation,
- Each control affects multiple bones in a coordinated and intuitive manner, e.g., lip spread, rather than controlling highly localized position on the skin. This allows a single animation control to affect multiple bones with a more coordinated and predictable result for the animator.

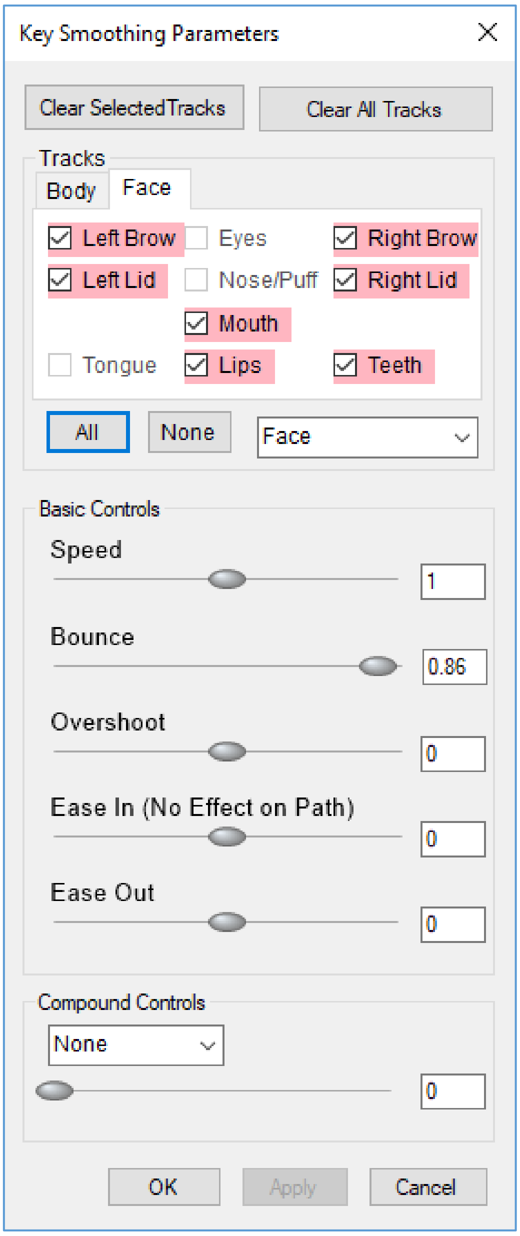
To compliment the Expression Builder, the Paula system provides an interface (Figure 5) offering the following animation parameters for each facial muscle control group. They are based on a Tension-Bias-Continuity interpolator (Bartels, Beatty, & Barsky, 1995), but with the parameters renamed to give a more intuitive set of controls for the animator:

- Speed: maps to the Tension parameter and controls the rate of change of the control values through the key
- Bounce: maps to the Continuity parameter and controls the degree to which the speed changes abruptly or more smoothly at a key
- Overshoot: controls the degree and direction of overshoot through a key which is an inherent feature of most interpolators and can be beneficial for creating abrupt “snap” effects
- Ease In/Out: Controls the classical Disney style animation features of ease at a key
- Compound controls: These are special controls that coordinate settings of the other controls for specific effects

These controls afford the animator with more direct ways to create the explosive motion out of a B or a P, or to create softer entries more subtle motions into a pose.

7. Conclusion and future work

We have retooled all the controls in the Expression Builder to this scripted approach. It gives artists more flexibility in creating not only lip postures, but convincing poses involving the entire face.


Key Smoothing Parameters

Clear SelectedTracks | Clear All Tracks

Tracks: Body | Face

☑ Left Brow □ Eyes ☑ Right Brow
☑ Left Lid □ Nose/Puff ☑ Right Lid
☑ Mouth
□ Tongue ☑ Lips ☑ Teeth

All | None | Face

Basic Controls
Speed: 1
Bounce: 0.86
Overshoot: 0
Ease In (No Effect on Path): 0
Ease Out: 0

Compound Controls
None: 0

OK | Apply | Cancel

Figure 5: Animation controls.

The approach has also lightened data storage demands. The previous version of the Expression Builder required over 800 values to record a single facial pose. Now the Expression Builder only stores the 56 slider values from the user interface, but the slimmed-down value set allows for more precision and flexibility in creating lip postures and eye apertures. Additionally, the more compact representation has made it possible to control the local speed of the motion of a slider at each key using Tension-Continuity-Bias controls that makes it easy to control overshoot, bounce, and many other dynamical properties. Future plans are to create additional viseme sets to support mouthing in multiple signed languages, including LSF, GSL, DGS and DSGS, and then to test the resulting animations with the Deaf community.

8. Acknowledgements

Many thanks to Ben Sturr, Anthony Bonzani, Nicole Barnekow, Syd Klinghoffer, and Andrew Alexander, the incredible team of modelers and animators of The American Sign Language Avatar Project at DePaul University for their dedication, enthusiasm, and good humor in creating fabulous animations with buggy software.

9. Bibliographic References

Alexa, M. (2002). Recent advances in mesh morphing. *Computer Graphics Forum, 21*(2), 173-198.

Austrian Association of Applied Linguistics. (2019, August). *Position Paper on Automated Translations and Signing Avatars.* Récupéré sur verbal; Verband für Angewandte Linguistik Österreich: https://www.verbal.at/stellungnahmen/Position_Paper-Avatars_verbal_2019.pdf

Baker, C., & Cokely, D. (1980). American sign language. *A Teacher's Resource Text on Grammar and Culture. Silver Spring, MD: TJ Publ.*

Baker-Shenk, C. (1983). A microanalysis of the nonmanual components of questions in American Sign Language.

Bartels, R. H., Beatty, J. C., & Barsky, B. A. (1995). *An introduction to splines for use in computer graphics and geometric modeling.* Los Altos, CA: Morgan Kaufmann.

Brumm, M., Johnson, R., Hanke, T., Grigat, R.-R., & Wolfe, R. (2019). Use of avatar technology for automatic mouth gesture recognition. *SignNonmanuals 2*.

Chen, S., Lou, H., Guo, L., Rong, Q., Liu, Y., & Xu, T.-M. (2012). 3-D finite element modelling of facial soft tissue and preliminary application in orthodontics. *Computer methods in biomechanics and biomedical engineering, 15*(3), 255-261.

Ebling, S., Wolfe, R., Schnepp, J., Baowidan, S., McDonald, J., Moncrief, R., . . . Tissi, K. (2015). Synthesizing the finger alphabet of Swiss German Sign Language and evaluating the comprehensibility of the resulting animations. *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, 10-16.

Jain, P., & Rathee, M. (2021). *Anatomy, Head and Neck, Orbicularis Oris Muscle.* Treasure Island, Florida: StatPearls Publishing.

Jennings, V., Elliott, R., & Kennaway, R. (2010). Requirements for a signing avatar. *Workshop on Copora and Sign Language Technologies (CSLT), LREC*, (pp. 133-136). Malta.

Johnson, R. (2022). Improved facial realism through an enhanced representation of anatomical behavior in signing avatars (submitted). *Seventh Sign Language Translation and Avatar Technology Workshop, Language resources and Evaluation Conference.* Marseilles: ELRA.

Johnson, R., Brumm, M., & Wolfe, R. (2018). An Improved Avatar for Automatic Mouth Gesture Recognition. *Language Resources and Evaluation Conference* (pp. 107-108). Myazaki, Japan: European Language Resources (ELRA).

Kipp, M., Heloir, A., & Nguyen, Q. (2011). Sign language avatars: Animation and comprehensibility. *International Workshop on Intelligent Virtual Agents*, (pp. 113–126).

Kipp, M., Nguyen, Q., Heloir, A., & Matthes, S. (2011). Assessing the deaf user perspective on sign language avatars. *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility. ACM*, (pp. 107-114).

Nadolske, M. A., & Rosenstock, R. (2007). Occurrence of mouthings in American Sign Language: A preliminary study. *Visible variation: Comparative studies on sign language structure*, 35–61.

Ostermann, J. (2002). Face Animation in MPEG-4. *MPEG-4 Facial Animation: The Standard, Implementation And Applications*, 17-55.

Parent, R., King, S., Fujimura, & Osamu. (2002). Issues with lip sync animation: can you read my lips? *Computer Animation 2002 (CA 2002)* (pp. 3-10). Geneva, Switzerland: IEEE.

Parke, F. I., & Waters, K. (2008). *Computer facial animation.* Boca Raton, Florida: CRC Press.

Piekartz, H. v., & Mohr, G. (2014). Reduction of head and face pain by challenging lateralization and basic emotions: a proposal for future assessment and rehabilitation strategies. *Journal of Manual & Manipulative Therapy, 22*, 24–35.

Reilly, I., & Anderson, D. (2002). FACES: The aquisition of non-manual morphology in ASL. *Directions in sign language acquisition , 2*, 159--182.

Schnepp, J., Wolfe, R., McDonald, J., & Toro, J. (2013). Generating Co-occurring Facial Nonmanual Signals in Synthesized American Sign Language. *Eighth International Conference on Computer Graphics Theory and Applications GRAPP/IVAPP*, (pp. 407-416). Barcelona.

Verlinden, M., Tijsseling, C., & Frowein, H. (2001). A Signing Avatar on the WWW. *International Gesture Workshop*, (pp. 169–172).

Wilbur, R. (1994). Eyeblinks & ASL phrase structure. *Sign Language Studies, 84*, 221–240.

Wolfe, R., Hanke, T., Langer, G., Jahn, E., Worsek, S., Bleicken, J., . . . Johnson, R. (2018). Exploring Localization for Mouthings in Sign Language Avatars. *Language Resources and Evaluation Conference* (pp. 207-212). Myazaki, Japan: European Language Resources Association (ELRA).