**README for the rigs “Angela\_noJALI.ma” and “Angela\_JALI.ma”**

**For use with the accompanying paper**

**“JALI: An Animator-Centric Viseme Model for Expressive Lip-Synchronization”**

**by Pif Edwards, Chris Landreth, Karan Singh and Eugene Fiume**

**14 April 2016**

We provide two Maya 2015 rigs which demonstrate JALI control as used in the procedural speech animation approach described in the accompanying paper.

**Angela\_noJALI.ma**

The character in this file is a conventional FACS-based rig using blendshapes to drive facial muscle actions, and skeletal joints to drive head, neck, jaw, tongue and eyelid actions. This rig uses controls labeled **CNT\_UPFACE, CNT\_MIDFACE** and **CNT\_LOFACE** to drive individual low-level facial poses, including blendshape-driven FACS muscle actions (pucker, sneer, frown, etc.) and joint-driven jaw actions (jawUpDn, JawInOut, etc.).

For speech, the rig uses controls labeled **CNT\_PHONEME** and **CNT\_COARTIC** to drive the character’s phonemes. Like the aforementioned low-level controls, these phoneme controls drive both blendshape and joint-driven actions on the face, but create unique visemes by driving appropriate combinations of these low-level facial, jaw and tongue actions.

The control **CNT\_PHONEME** drives pose-to-pose-animated visemes (AHH, OHH, SSS, etc.). Animators will typically key a single phoneme to a non-zero value, for a given time when that phoneme is uttered in the actor’s voice track. All other phonemes will typically be keyed at zero for that given time.

The control **CNT\_COARTIC** is used for visemes that co-articulate with those driven by **CNT\_PHONEME**. For example, when pronouncing “S” as in “Sue”, the “S” viseme must show lip pursing from the ensuing “U” phoneme, which would not be present when pronouncing “S” as in “See”. The channel “WA\_PEDAL” in **CNT\_COARTIC** allows pursing the character’s lips (i.e., “U”) with timing that is independent of the timings of the surrounding phonemes (i.e., “S”). Other channels in **CNT\_COARTIC** allow co-articulating labial (MMM), labiodental (FFF) and tongue-only (LNTD, GK) phonemes with adjacent phonemes keyed with the **CNT\_PHONEMES** control.

**Angela\_JALI.ma**

The character in this file is identical to that of the conventional rig, but has an additional control, **CNT\_JaLi**. This control sets the relative level of speech governed by jaw/tongue action (“Jaw”, **CNT\_JaLi**.xT) and speech governed by facial muscle action (“Lip”, **CNT\_JaLi**.yT), as described in Section 3 of the accompanying paper. To demonstrate this, please select a phoneme in **CNT\_PHONEME** and set it to its maximum value of 12.0. Then select **CNT\_JaLi** and move it around its local XY space. You will clearly see that for that phoneme, the character will pose her face in a wide variety of speaking styles, as shown in Figure 2 of the accompanying paper.

**Rigging a JALI Face**

The JALI rig uses an enhanced network of utility, expression and function curve (i.e., “set-driven-key”) nodes to provide independent Jaw and Lip control to the character’s facial appearance. Figure (X) shows a comparative view of this network in Maya’s dependency graph window, between a conventional facial rig and a JALI rig. This view shows the nodes which are activated when the character pronounces the phoneme “R” (**CNT\_PHONEME**.RRR = 10.0).

In the conventional rig (Figure Xa), the attribute **CNT\_PHONEME**.RRR is connected to an appropriate combination of FACS-based blendshapes (pucker, loLipOut, upLipOut), jaw rotation (jaw.rX) and tongue position (tongue.rX, tongue.tY and tongue.tZ). The resulting values on these blendshape, jaw and tongue positions are determined by set-driven-key function curves **“FC\_\*\_RRR”,** which are combined through blend nodes **“BN\_\*”** with values from other inputs (e.g**. CNT\_LOFACE**, not shown). The function curves map the input **CNT\_PHONEME**.RRR to appropriate blendshape/jaw/tongue poses that create an “R” appearance on the character’s face (Figure Xc).

In the JALI rig (Figure Xb), the attribute **CNT\_PHONEME**.RRR is connected to these same blendshape/jaw/tongue nodes, but is filtered through additional nodes which modify the “RRR” output using the additional **CNT\_JaLi** control. The Jaw and Lip values on **CNT\_JaLi** are multiplied with the **CNT\_PHONEME**.RRR output through the nodes “**MULT\_\*\_RRR**”, thus scaling the resulting values on the blendshapes (pucker, etc.), jaw and tongue positions on the character’s face.

In the example of Figure X, “Jaw” is set relatively low (2.0) and “Lip” is set relatively high (14.0)--as in Figure 2 of the accompanying paper, she is pronouncing the R phoneme “through her teeth.” The **MULT\_LIP\_RRR** node correspondingly scales the FACS blendshape values (pucker, etc.) to a higher value than in the conventional rig, and the **MULT\_JAW\_RRR** and **MULT\_TNG\_RRR** nodes scale the jaw/tongue values to a lower value than the conventional rig. The resulting appearance on the face (Figure Xd) shows the character pronouncing “R” with enhanced FACS lip action, and diminished jaw/tongue action, compared to the conventional rig.

