The CASSI System for Animated Speech Simulation

November Scheidt and Howard J. Hamilton

Technical Report CS-99-01
Mar., 1999

Department of Computer Science
University of Regina
Regina, Saskatchewan, CANADA
S4S 0A2

ISSN 0828-3494
ISBN 0-7731-0378-3

The CASSI System for Animated Speech Simulation

November Scheidt and Howard J. Hamilton

Department of Computer Science, University of Regina

Regina, Saskatchewan, Canada S4S 0A2

E-mail: {nova,hamilton}@cs.uregina.ca

Abstract

We describe the CASSI (Computer Animated Speech SImulator) software system, which creates facial animations corresponding to X-ray microbeam data. Each data file records a sequence of positions of the tongue, lips, and jaw of a human subject. The CASSI system utilizes Parke's model of a human head, augmented by a representation of the inside of the mouth, including the tongue. CASSI's output is a 3D facial animation that shows, in animated form, a subject's facial movements while speaking and performing other tasks. The output animations are subjectively evaluated with respect to the positions of the teeth, lip, and tongue for a variety of human subjects and tasks.

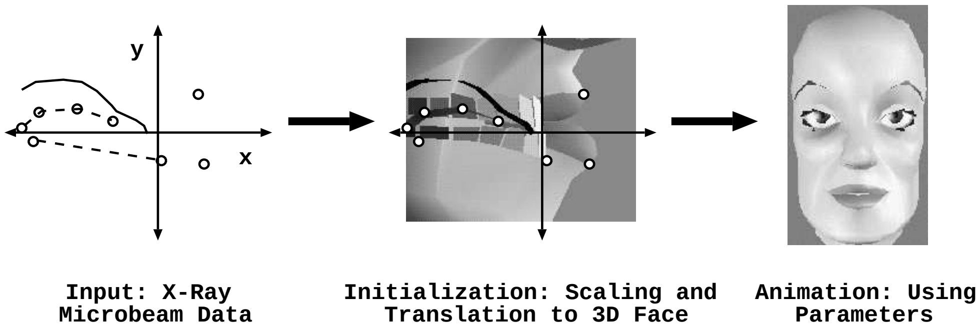


Figure 1: Overview of Animation

1 Introduction

We describe the CASSI (Computer Animated Speech SImulator) software system, version 1.0, June 1998, which creates facial animations corresponding to X-ray microbeam data. Each data file records a sequence of positions of the tongue, lips, and jaw of a human subject. This system utilizes a modified version of Parke's model [4] of a human head. CASSI augments Parke's model by representing the inside of the mouth, including the tongue, as has also done in previous research[1, 3, 2]. CASSI's output is a 3D facial animation that shows, in animated form, a subject's facial movements while speaking and performing other tasks.

The remainder of this report is organized as follows. In Section 2, an overview of the CASSI software system is presented. In Section 3, we explain how the facial model is initialized and how it is animated, including a detailed explanation of our model of jaw rotation. Then in Section 4, we present a subjective evaluation based on a detailed examination of the appearance of the teeth, lips, and tongue of the animated model for a variety of human subjects and tasks. Finally, in Section 5, we draw conclusions and make suggestions for further research.

2 Overview

Figure 1 shows an overview of the CASSI system. The input is based on X-ray microbeam data measuring the position of the lips, tongue, and jaw. During the initialization phase, the 3D facial model is adjusted, using translation and scaling, to provide a starting position for the animation. During the animation phase, the parameters of the facial model are changed corresponding to the movement of the lips, tongue, and jaw;

these changes in parameter values cause corresponding changes in the facial model, as it is displayed frame by frame. The input, initialization phase, and animation phase are now briefly described.

The input to the system is X-ray microbeam data obtained from the University of Wisconsin [5]. The main part of the data consists of a series of data files, each describing one subject performing one task. Each data file contains a time series of the x-y coordinates of eight pellets (shown as open circles in Figure 1). The x-y coordinates represent a side view of the mouth region of a human head, with the origin defined as the bottom of the top front teeth. The palatal outline (or roof of the mouth) is shown as the solid curving line in Figure 1. The eight pellets include: one for the upper lip and one for the lower lip (shown as the right-most open circles), two points on the jaw (shown as the lowest connected open circles), and four points along the tongue (shown as four points connected in a line directly beneath the palatal outline).

Currently, we have access to approximately 1700 data files, representing 15 human subjects performing up to 118 different tasks. (The Wisconsin data included a 16th subject, JW29; we did not use data for JW29 because all data files had errors of various types, especially due to incomplete recording of the location of the back jaw pellet, which is required for our method.) Some tasks are not recorded for every speaker; as well, some data files contain invalid data due to problems (such as mistracking pellet positions) during data collection. In addition, we have access to a file called *Pal.dat*, which describes the palatal outline for each subject, and a file called *Headmeasures.txt*, which gives ten measurements of each subject's head. We always refer to these latter two files by name, and we do not include them in the general term *data files*.

Each line in a data file contains a timestamp and the x and y coordinates for the 8 pellets at that time. The changes in the positions of the pellets correspond to the changes in the positions of the lip, jaw, and tongue as the speaker performs a task. The tasks include: (1) saying words, sentences, VCV (vowel consonant vowel) patterns, and paragraphs, (2) swallowing, (3) maximum tongue protrusion, and (4) jaw wagging. Movement in the positive x direction is towards the front of the face, and movement in the positive y direction is towards the top of the head.

For the initialization phase, the face is adjusted to a starting location. The starting location is assumed to be the first valid line in the data file. In all data files, the first few lines have pellet positions that are "out of range" (1000000), indicating that the pellets are not yet being tracked. The first valid line is the first one

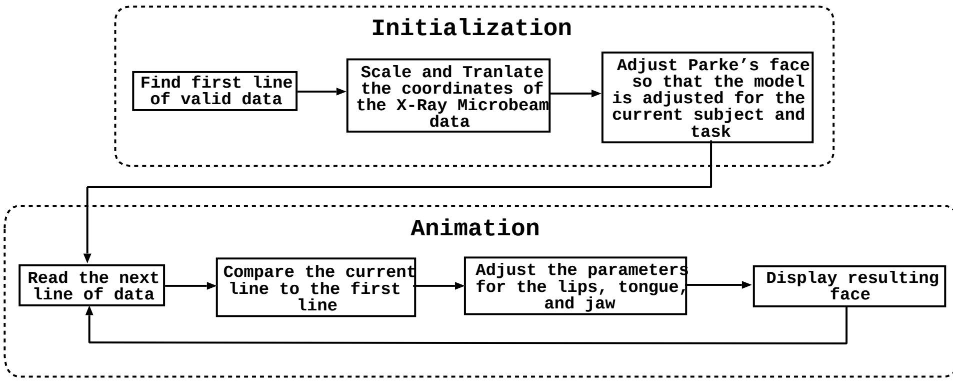


Figure 2: The Initialization and Animation Phases

that has no "out of range" value. Once a valid line has been found, scaling and translation are applied to the pellet coordinates to adjust them relative to the 3D facial model. From these relative pellet locations, the facial model's lips, tongue and jaw are adjusted. This means that the lips, tongue, and jaw (in particular) are set to initial values that depend on the particular subject and task that the data file represents.

For animation to occur, each data line containing the current pellet positions is read and the parameters of the facial model are adjusted to represent the shift in the current pellet position from the initial positions. The animation represents the movement of the actual human subject based on the motion of the pellets placed on the tongue, lips, and jaw. Depending on the speed of the graphical processing, some of the lines of input are ignored. In our current work on a SGI Indigo with a 150 MHZ R4400 CPU, we use every 10th line of the data. On faster machines, such as a SGI O2 with a 200MHZ R5000 CPU, every 5th line of data can be displayed.

3 Initialization and Animation

Figure 2 illustrates, in further detail, the initialization and animation phases. In the initialization phase, there are three major steps. The first step scans through the X-ray microbeam data file for the first line of valid (no "out of range") data. The second step maps the pellet positions to the 3D facial model. This mapping is done through scaling and translation. The scaling is based on the ratio between the jaw lengths of the animated model and the human subject. More precisely, each coordinate is scaled by multiplying by the ratio p/d, where d is the distance between the two pellets on the jaw of the live subject, and p is the distance between two approximately corresponding points in the model. After scaling, translation is done so that the origin (for the pellet positions) is translated to the point corresponding to the bottom of the top front teeth in the facial model. The pellet positions now have a location relative to the 3D facial model. Lastly, the third step adjusts the 3D facial model to match these pellet postions. The tongue is positioned according to the four tongue pellet positions, the upper and lower lips are positioned according to their associated pellets placed on the vermillion border (the upper and lower edges of the lips respectively), and the lower teeth are positioned according to the two pellets on the jaw using jaw rotation, as discussed in Section 3.1. To compensate for high lips, the tip of the nose is also adjusted so that there is always a space between the upper lip and the nose.

During the initialization step, the palatal outline (or roof of the mouth) is drawn, using additional data from the *Pal.dat* file. Each subject has an individual palatal outline, described as a series of x and y coordinates in *Pal.dat*, representing the palatal centerline. The number of x and y coordinate pairs varies among the subjects, because points were included in *Pal.dat* as needed to show the palatal outline. We implemented a three dimensional model of the palate based on the two dimensions provided in the *Pal.dat* file and assumptions about the probable width and shape of the palate. The palate is shown as the darkest object in Figure 3, where the tongue is shown immediately underneath in light grey. The palate is narrow at the front of the mouth and wider toward the back. It has a slight downward slope from its centerline to its edges.

Also, during the initialization step, the shapes of the jaw and chin are adjusted according to the measured positions of the gonion and gnathion, as given in the *Headmeasures.txt* file. Both the gonion and the gnathion are part of the jaw bone. The *gonion* is defined as the point at the angle of the jaw (in the back of the jaw, where the jaw angles upward towards the ears). The *gnathion* is the point of the chin. The shape of the jaw is adjusted using the jaw rotation formulas described in Section 3.1.

After the initialization phase is complete, the animation phase occurs in four major steps (see Figure 2). The first step reads the next line of data, containing the current x and y coordinates for the eight pellets. The second step computes the movement (displacement) of each of the current pellet positions from the initial pellet positions (as recorded in the initialization phase). Each displacement is scaled (using p/d) to

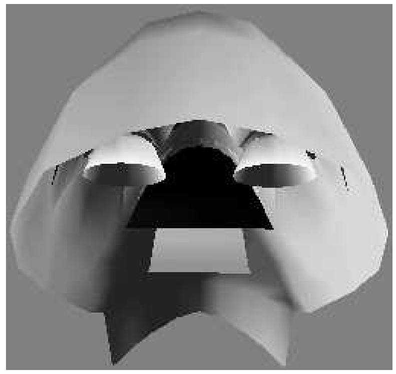


Figure 3: Top View of the Palatal Outline, as seen from the backside of the face

Parameter #	Initial Value	Description
55	0.0	T1 X movement
56	0.0	T1 Z movement
57	0.0	T2 X movement
58	0.0	T2 Z movement
59	0.0	T3 X movement
60	0.0	T3 Z movement
61	0.0	T4 X movement
62	0.0	T4 Z movement
63	0.0	Upper Lip X movement
64	0.0	Upper Lip Z movement
65	0.0	Lower Lip X movement
66	0.0	Lower Lip Z movement

Table 1: New Parameters

find the displacement relative to the initial facial model. In the third step, these scaled displacements are used to adjust the parameters of the lips, tongue, and jaw. Table 1 shows parameters added for lip and tongue movement. The first column lists the parameter number, the second column lists its initial value, and the third column describes the parameter (where the pellets are numbered in sequence T1, T2, T3, and T4, from front to back). Jaw movement is obtained using the jaw rotation formulas discussed in Section 3.1. In the fourth step, the face is displayed with the changes made to the face using the new parameter values. This animation phase is repeated for each line in the data file.

3.1 Jaw Rotation

The underlying method for adjusting the jaw in both initialization and animation is the same. The method is based on Parke's [4] original formula for jaw rotation, which described the rotation around a single fixed point, corresponding approximately to the back of the jaw. In our implementation, this formula is used to derive a method of calculating the point of rotation for each frame of animation. The approach used (shown in Figure 4) has two major steps: calculating the angle of rotation and finding the point of rotation. In

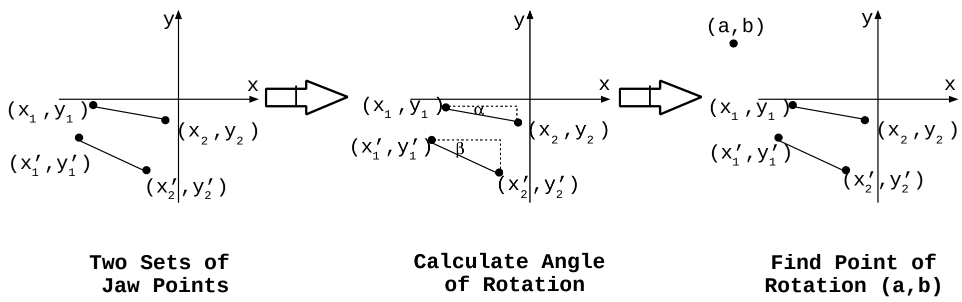


Figure 4: Rotation around the Origin

Figure 4, (x_1, y_1) and (x_2, y_2) are the original jaw points and (x'_1, y'_1) and (x'_2, y'_2) are derived by rotating line $(x_1, y_1), (x_2, y_2)$ around a fixed point. To calculate the angle of rotation, the angle of incline of the original jaw position (α) and of the current jaw position (β) must be found. These angles are obtained using Equations 1 and 2:

$$tan\alpha = \frac{y_1 - y_2}{x_1 - x_2} \tag{1}$$

$$tan\beta = \frac{y'_1 - y'_2}{x'_1 - x'_2} \tag{2}$$

Using these two angles, the angle of rotation(θ) is determined:

$$\theta = \alpha - \beta \tag{3}$$

The next step is to find the point of rotation (a,b). We begin with variations of the equations in Parke's model:

$$x' - a = (x - a)cos\theta + (y - b)sin\theta \tag{4}$$

$$y' - b = -(x - a)sin\theta + (y - b)cos\theta \tag{5}$$

From these, we derive the following equations for the point of rotation:

$$b = \frac{(y' + xsin\theta - ycos\theta)(1 - cos\theta) - (x' - xcos\theta - ysin\theta)(sin\theta)}{2 - 2cos\theta} \tag{6}$$

$$a = \frac{y' + xsin\theta - ycos\theta - b(1 - cos\theta)}{sin\theta} \tag{7}$$

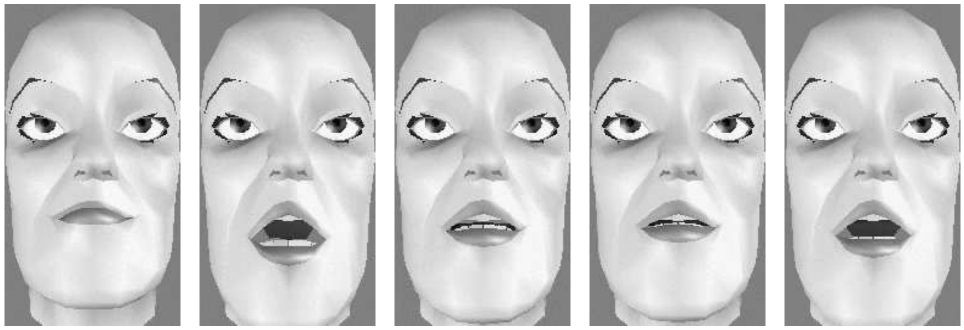


Figure 5: Front View of Jaw Wagging Task for JW27

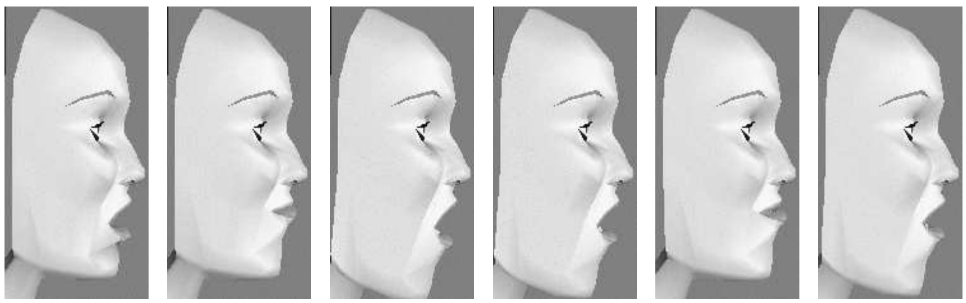


Figure 6: Side View of Jaw Wagging Task for JW27

In Equations 6 and 7, only one point (x_1, y_1) and its derived point (x'_1, y'_1) need to be used for (x, y) and (x', y'), respectively.

Once the angle of rotation and the point of rotation have been calculated, the teeth, jaw, cheeks, and neck can be moved in varying amounts similar to Parke's implementation. For instance in Parke's implementation, the surface jaw/chin points had 100% motion, the teeth had 90% motion, the points directly under the chin had 75% motion, the points in the upper neck had 50% motion, and some points in the cheek had 35% motion. To stay close to Parke's original model yet reflect the fact that the teeth should have 100% motion (corresponding to the actual pellet positions), this implementation multiplied all percentages by 1.11111 so that there was 111%, 100%, 83%, 56%, and 39% motion respectively.

The only difference between the initialization and the animation phase is the choice of (x_1, y_1),(x_2, y_2) (starting jaw location), and (x'_1, y'_1), (x'_2, y'_2) (rotated jaw). For instance, in the initialization phase, the starting jaw and teeth locations are the original placement of the points (Parke's original topology), and the rotated jaw and teeth points are from the placement of the jaw in *Headmeasures.txt* and the location of the teeth in the first line of valid X-ray microbeam data, respectively. in the animation phase, the starting jaw location is the jaw points from the first valid line of X-ray microbeam data, and the rotated jaw points are from each consecutive line of data.

Figure 5 and Figure 6 provide front and side view, respectively, of task #106 (the jaw wagging task) as performed by subject JW27. These views were taken at slightly different times, but both demonstrate that the jaw and lips move up and down together, even though their parameters are adjusted separately. In Figure 5, the tongue is visible as a very dark object inside the mouth, in the frames where the mouth is open.

4 Results: A Subjective Evaluation of System

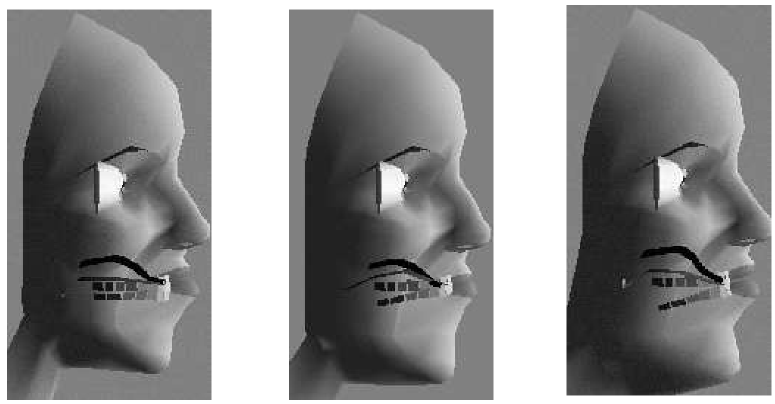
To evaluate the current CASSI system, the output facial animations for each subject were viewed from a variety of positions. Each facial animation was viewed from the front (viewing the full face from the starting orientation), the side (viewing the full face from a side orientation), the inside (viewing half of the face from a side orientation), and the top (viewing the top of the palatal outline). During the evaluation, parameters were not manually adjusted.

Table 2 shows the results of the evaluation for 15 human subjects (JW11 to JW502). Each subject was examined on at least two different tasks, as listed in the table. For instance, for JW16, task #3 was to read number sequences (e.g., "9739286 8495571 5945341") as individual digits with pauses between. For each subject, the tasks were arbitrarily selected from among the error-free files in the X-ray microbeam data. The positions of the teeth, lips, and tongue were subjectively evaluated, and a rating of EXCELLENT, VERY GOOD, GOOD, or FAIR was assigned to each component (teeth, lip, or tongue). The rating is based on the number of negative points (–'s) identified for that component: EXCELLENT means that there was nothing wrong with that component, VERY GOOD means that one problem was seen for that component (one –), GOOD means that two problems were seen (two –'s), and FAIR means that three problems were seen (three –'s). The following paragraphs provide more details about this rating system.

For the evaluation of the teeth positions, the front and side views were examined. The front view was examined to see if the teeth ran into the lips, which was considered undesirable. The main focus of evaluaton, however, was the side view looking inside the mouth. From this view, close attention was paid

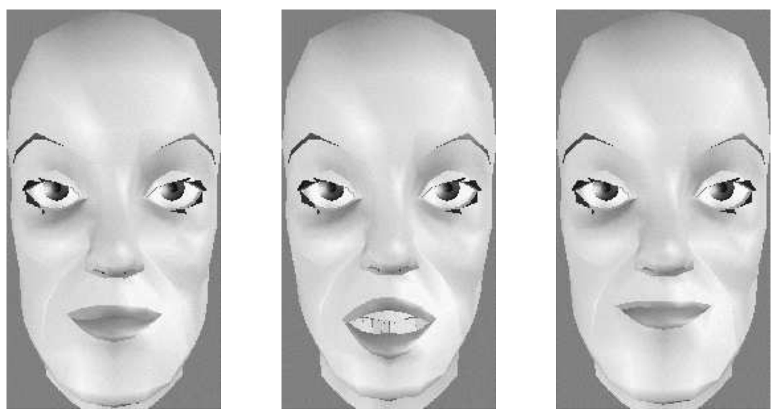
Subject	Task #	Description	Teeth Position	Lip Position	Tongue Position
JW11	19 21 106	sentences citation words jaw wagging	GOOD –from front view, bottom teeth run into lower lip at times –back teeth overlap (top to bottom)	FAIR – –there is a large separation between lips(they never touch) –lips are often below the bottom of the lower front teeth	FAIR – – –tongue pokes noticeably through upper and lower teeth
JW12	2 5	citation words citation words	GOOD –back teeth overlap(top to bottom) –bottom teeth slant downwards towards front teeth	VERY GOOD –lips go below the bottom of the lower front teeth	VERY GOOD –in task 2, tongue pokes through upper front teeth
JW15	9 10	citation words sentences	GOOD –front teeth overlap (top to bottom) –bottom teeth slant upward towards front teeth	GOOD –lips are tightly placed together –lips go below the bottom of the lower front teeth	GOOD –the tongue pokes through the upper front teeth (more noticeable in task 10) –from the front view, the tongue is barely viewable (hidden by teeth)
JW16	3 6	number sequences citation words	VERY GOOD –bottom teeth slant upward toward front teeth	GOOD –there is a separation between lips –top lip seems too high	VERY GOOD –tongue pokes through top front teeth only in task 3
JW18	4 14	citation words citation vowels	FAIR –front teeth overlap (top to bottom) – –bottom teeth slant upward toward front teeth	VERY GOOD –lips are tightly placed together	GOOD – –for task 14, tongue pokes slightly through upper and lower teeth
JW19	1 23	citation words citation words	VERY GOOD –bottom teeth slant upward toward front teeth	GOOD – –there is a separation between lips	GOOD – –tongue pokes through upper and lower teeth (more noticeable in task 23)
JW21	7 12	sentences paragraph	GOOD –from front view, bottom teeth run into lower lip and (in task 12) the upper lip –teeth overlap (top to bottom) +bottom teeth horizontally aligned	EXCELLENT	FAIR – – –tongue pokes noticeably through upper and lower teeth
JW24	17 18	sentences citation words	EXCELLENT +bottom teeth horizontally aligned	FAIR – –from side view, lower lip protrudes out farther then upper lip –lips go below the bottom of the lower front teeth +from front view,lips are nicely closed	FAIR – – –tongue pokes noticeably through upper and lower teeth
JW25	15 16	vowel sequences citation VCV	VERY GOOD –bottom teeth slant upward toward front teeth(slight)	GOOD –lips are tightly placed together (especially in task 16) –from side view, lower lip protrudes out farther than upper lip	VERY GOOD –In task 16, tongue pokes through lower teeth
JW27	20 22 106	sentences citation words jaw wagging	GOOD –front teeth overlap (top to bottom) –bottom teeth slant upward toward front teeth	VERY GOOD –lips are tightly placed together	GOOD – – tongue pokes through upper and lower teeth
JW32	13 15	citation sVd's eg. side vowel sequences	GOOD –from front view, teeth run into upper lip (especially in task 13) –teeth overlap (top to bottom) +bottom teeth horizontally aligned	VERY GOOD –from side view, lower lip protrudes out farther than upper lip (slight) +lips are nicely closed	EXCELLENT
JW40	5 8	citation words citation words	VERY GOOD –bottom and top teeth seem too far apart +teeth are evenly separated (top to bottom)	VERY GOOD –there is a separation between lips	VERY GOOD –in task 8, tongue pokes through upper teeth
JW41	1 11	citation words paragraph	FAIR – – –bottom teeth slant upward toward front teeth	VERY GOOD –lips go below the bottom of the lower front teeth	GOOD – –tongue pokes through upper teeth for both tasks and the lower teeth in task 11
JW45	3 13	number sequences citation sVd's eg. side	VERY GOOD –bottom teeth slant upward toward front teeth (slight)	EXCELLENT	GOOD –in task 13, the tongue pokes through the upper front teeth –from front view, tongue is barely viewable (seems too high above the top teeth)
JW502	21 22	citation words citation words	VERY GOOD –bottom teeth slant upward toward front teeth (slight)	VERY GOOD –lips are tightly placed together	EXCELLENT

Table 2: Subjective Evaluation of Subjects and Tasks



JW24 **JW27** **JW41**

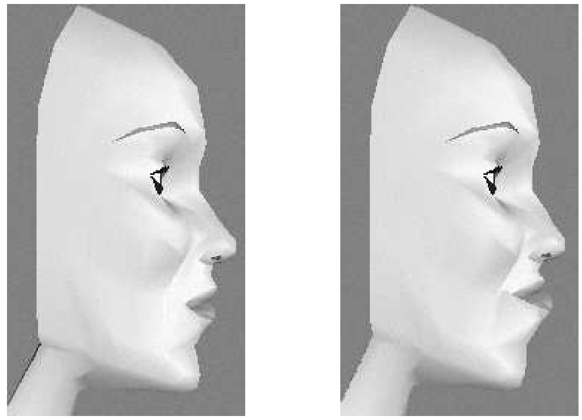
Figure 7: Internal side view of EXCELLENT (JW24), GOOD (JW27), and FAIR (JW41) teeth positioning



JW45 **JW11** **JW502**

Figure 8: Front view of EXCELLENT (JW45), FAIR (JW11), and VERY GOOD (JW502) lip positioning

to the positioning of the lower set of teeth while the task was being performed, and particularly at the end of the task. Subject JW24 was considered to have ideal positioning for the teeth because the bottom teeth were horizontally aligned with a very slight overlap in the front teeth (see Figure 7). All other subjects were compared to JW24. For instance, when compared to JW24, JW27 (in Figure 7) had front teeth that overlapped from top to bottom (this was one –) and lower teeth that sloped upward from back to front (this was another –). These two factors (two –'s) caused JW27 to have an overall rating of GOOD for teeth position. Another example is JW41 who had a very steep slope upward from back to front, when compared to the other subjects (see Figure 7). Since this slope was extremely large, three negatives were assigned to it; this gave JW41 an overall rating of FAIR for teeth position.



JW45 JW24

Figure 9: Side view of EXCELLENT (JW45) and FAIR (JW24) lip positioning

For the evaluation of the lip positions, the front and side views were used. The front view (in the original starting location) (see Figure 8) was examined to see if the lips lightly touched (desirable), were too far apart (as in JW11), or were too close together (as in JW502). The front view was also examined to see if the lips went above the top of the upper front teeth or below the bottom of the lower front teeth. The side view (see Figure 9) was examined to see if the bottom protruded more than the top (as in JW24). Again, for situations with significant problems, extra negatives were assigned. For instance, when compared to others, JW11 (in Figure 8) had lips that were widely separated and, in fact, never touched. For this large lip separation, JW11 was assigned two –'s. JW11 also had lips that went below the bottom of the lower front teeth (one –), which gave JW11 an overall rating of FAIR (with three –'s).

The tongue position was evaluated only from the front view. The front view was examined (from the original starting location) to see if the tongue poked through the teeth (see Figure 10). If the tongue poked through either the upper or the lower front teeth (as in JW12), then one negative was assigned. If the tongue poked through both the upper and lower front teeth (as in JW27), then two negatives were assigned. If the tongue poked very noticeably through the upper and lower front teeth (as in JW21), then three negatives were assigned. Also from the front view, one negative was assigned if the tongue was barely viewable (as in JW15). JW502 and JW32 were rated as EXCELLENT because, from the front view, the tongue was visible and did not appear to poke through the teeth.

The tongue was not evaluated with regard to its shape from the internal side view or whether it poked

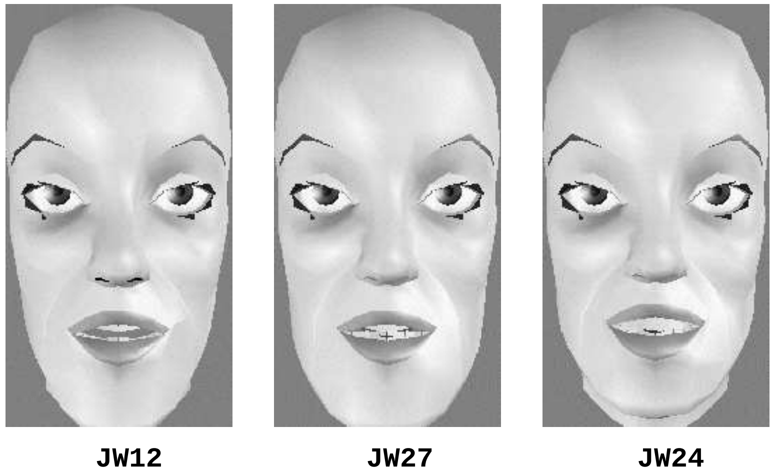
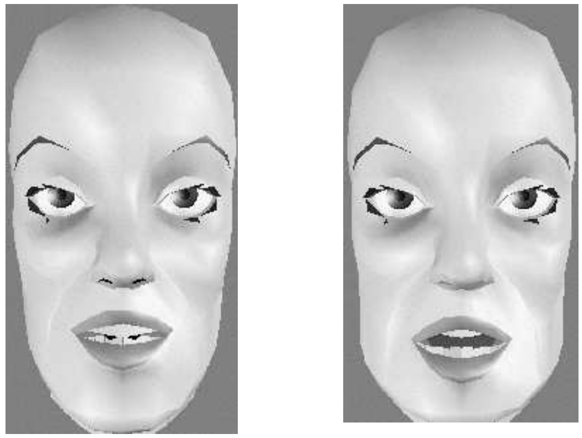


Figure 10: Front view of tongue poking through the teeth for JW12 (VERY GOOD) JW27 (GOOD), and JW24 (FAIR)

through the palatal outline. From the internal side view, the tongues for all tasks for all subjects (with the exception of task #5 for JW40) had a V-shape in the back at times during the animation (this V-shape can be seen in subject JW41 in Figure 7). This V-shape results from the initialization of the farthest back points in the tongue. When the tongue is initialized, the farthest back points are "anchored" to a position that is farther back and down from the tongue's back pellet position (T4). This anchoring might be improved in one of three ways: by eliminating the anchored points, by making the anchored points farther back and down, or by making the anchored points move when the tongue moves. Also, the tongue was not evaluated on whether or not it poked through the palatal outline. In all but a few cases, the tongue did poke through the palatal outline (particularly the alveolar ridge). Since this behavior did not affect the external appearance of face, it was not assigned a negative point.

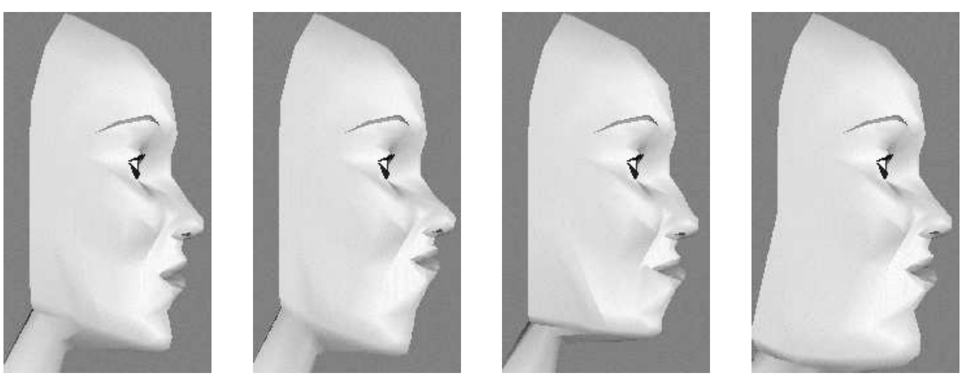
In overall appearance, a few subjects had some quirks. For instance, JW19 (and to a lesser extent JW32) had a palatal outline that extended below the tip of the front upper teeth (see Figure 11). This is because the palatal outline was defined to have thickness and to go down at the edges. A few subjects had unusually shaped faces (see Figure 12). For instance, JW18 had a very long jaw, JW32 had a very square jaw, and JW41 had an unusual shape to the back of his jaw.

Overall, in Table 2, the ratings for teeth position were 1 EXCELLENT, 6 VERY GOOD, 6 GOOD, and 2 FAIR, the ratings for lip position were 2 EXCELLENT, 7 VERY GOOD, 4 GOOD, and 2 FAIR, and the



JW19 JW32

Figure 11: Front view of JW19 and JW32 with palatal outline extending below the top front teeth



JW45 JW18 JW32 JW41

Figure 12: Side view of a nicely shaped jaw (JW45) and unusually shaped jaws (JW18, JW32, and JW41)

ratings for tongue position were 2 EXCELLENT, 4 VERY GOOD, 6 GOOD, and 3 FAIR. The subject who appears to be animated best is JW502, with ratings of VERY GOOD, VERY GOOD, and EXCELLENT. The subject who appears to animated worst is JW11, with ratings of FAIR, FAIR, and GOOD.

The variety of ratings awarded to the subjects suggests that certain faces fit the original mesh better than others. For example, with regard to lip positioning, some subjects had lips that were too close together, and other subjects had lips that were too far apart. Possibly some subjects have thicker or thinner lips, but the current implementation makes all lips the same size and shape. Thinner lips should be used for those whose lips are held tightly together, and thicker lips should be used for those whose lips never touch. For 13 out of 15 subjects in Table 2, the tongue poked through the front teeth for at least one task. Collision detection should be implemented to ensure that the tongue never passes through the teeth. The positioning of the teeth also seems to be a problem with some subjects, such as JW41 who is shown with a large slope in the lower teeth. This unusual placement should be verified. If it is correct, some adjustments might be made to improve the appearance, such as adjusting the length of the teeth and sloping the upper teeth to match the lower teeth.

5 Conclusions and Future Research

In this report, we described the CASSI software system for creating facial animations corresponding to X-ray microbeam data. We explained the system by describing the input, the initialization phase, and the animation phase. CASSI's output is a 3D facial animation that shows, in animated form, a subject's facial movements while speaking and performing other tasks. As has been done by previous researchers, we added a tongue to Parke's facial model. The major achievement of this research was making the animated tongue, teeth, jaw, and lips move according to X-ray beam data. To our knowledge, this is an original achievement. The output animations were subjectively evaluated with respect to the positions of the teeth, lip, and tongue for a variety of human subjects and tasks. The results of this evaluation suggest that CASSI provides excellent or very good animations about half the time and fair or good animations about half the time. The animations for certain subjects appeared better than others, possibly due to a better match between these subjects and the facial model. For instance JW502 appears to match the model well, whereas

JW11 does not.

Four new features that could be added to the CASSI system are speech synchronization, lip rounding, lip thickness, tongue thickness, and increased facial adjustment to the data. Speech sychronization should be implemented so that the corresponding sound is produced while the facial animation is displayed; this addition would provide further feedback on the movements of the lips, jaw, teeth, and tongue in relation to the speech sounds being produced. In the current implementation, from the front view, the corners of the lips do not move horizontally towards or away from each other. Lip rounding should be implemented so that the lips round (or the corners move together) when the lips are brought forward. The thickness of the lips should be adjusted for each speaker by using the upper and lower lip pellet positions in the first valid line of data, which likely corresponds to a gently closed mouth. In our current implementation, the tongue is a flat surface with no thickness. Thickness should be added to the tongue so that it can be seen better. Because subjects have differently shaped faces, the 3D facial model should be further adjusted to more closely match the subject. For instance, the width of the face should be adjusted according to data provided for each subject in the *Headmeasures.txt* file.

The problems noted in Section 4 should also be fixed. These include: (1) adding collision detection to ensure that the tongue does not poke through the teeth or the palatal outline; (2) adjusting the anchored points in the tongue so these points move along with the rest of the tongue, are eliminated, or are moved farther back and down; (3) fixing the palatal outline in subjects JW19 and JW32 so that it does not extend below the tip of the top front teeth; (4) verifying the position of the teeth by checking with the software provided with the X-ray microbeam data; and (5) verifying the position of the jaw by checking with the video tape of the subjects performing specific tasks.

Acknowledgement: We would like to thank Carl Johnson and the other members of the X-ray microbeam group at the University of Wisconsin for providing us with data, as supported by research grant number R01 DC 00820 from the National Institute on Deafness and Other Communicative Disorders, U.S. National Institutes of Health (X-ray Microbeam data). Our research was supported by the Natural and Engineering Research Council of Canada by means of a Postgraduate Scholarship (Scheidt) and a Research Grant (Hamilton).

References

[1] J. Beskow. Rule-based visual speech synthesis. In *Proceedings of Eurospeech '95*, Madrid,Spain, September 1995. Paper obtained from http://www.speech.kth.se/multimodal/papers/.

[2] M. M. Cohen and D. W. Massaro. Modeling coarticulation in synthetic visual speech. In N. Thalmann and D Thalmann, editors, *Models and Techniques in Computer Animation*, pages 141–155. Springer-Verlag, Tokyo, 1993. Paper obtained from http://mambo.ucsc.edu/psl/ca93.html.

[3] M.M. Cohen, J. Beskow, and D.W. Massaro. Recent developments in facial animation: an inside view. In *Auditory-Visual Speech Processing (AVSP)*, Terrigal, New South Wales, Australia, December 1998. Paper obtained from http://www.cmis.csiro.au/avsp98.

[4] F.I. Parke and K. Waters. *Computer Facial Animation.* A K Peters, Wellesley, MA, 1996. Code available as appendices at: http://www.crl.research.digital.com/publications/books/waters/waters_book.html.

[5] J. R. Westbury. *X-Ray Microbeam Speech Production Database User's Handbook.* Waisman Center, University of Wisconsin, Madison, WI, June 1994.