RUTH B. GROSSMAN AND JUDY KEGL

# To Capture a Face: A Novel Technique for the Analysis and Quantification of Facial Expressions in American Sign Language

Over the past two decades research on American Sign Language (ASL) has shown that, although the hands and arms articulate most of the content words (nouns, verbs, and adjectives), a large part of the grammar of ASL is expressed nonmanually. The hands and arms do play central grammatical roles, but, in addition, movements of the head, torso, and face are used to express certain

Ruth B. Grossman obtained her Ph.D from Boston University in neurolinguistics in 2001. She has been teaching at Boston University on adult neurogenic disorders and is a postdoctoral research fellow with the Lab for Developmental Cognitive Neuroscience at the Boston University Medical School. Her current field of investigation is the interaction of language and facial expressions in children with autism. Judy Shepard-Kegl is currently Full Professor in the linguistics program at the University of Southern Maine and Director of the Signed Language Research Laboratory there where she oversees several research projects on the crosslinguistic comparison of signed languages and the emergence of a new signed language in Nicaragua and coordinates the training of ASL/English interpreters under a grant to the university from the Maine Department of Education. She received her Ph.D. in linguistics from M.I.T. in 1985 and has completed extensive postdoctoral studies in neuroscience at Rutgers, The State University of New Jersey. She has published extensively on the linguistics of signed languages as well as spoken and signed language aphasia. Her research addresses a unique blend of theoretical and applied dimensions that both contribute to theoretical debates and are sensitive to humanitarian concerns.



Sign Language Studies Vol. 6 No. 3 Spring 2006

................15915$ $CH2 03-23-06 12:03:41 PS PAGE 273

aspects of ASL syntax such as functional categories, syntactic agreement, syntactic features, complementizers, and discourse markers. Since the pioneering work on nonmanuals (i.e., parts of ASL not expressed through the arms and hands) by Liddell (1986), Baker-Shenk (1983, 1986), and Baker-Shenk and Padden (1978), research has increasingly focused upon facial expressions in ASL and their syntactic significance (Neidle et al. 2000; Aarons 1994; Aarons et al. 1992; Baker-Shenk 1985).

It is now well established that ASL requires the use of the face not only to express emotions but also to mark several different kinds of questions: wh-questions (questions using *who, what, where, when,* or *why*), yes/no (y/n) questions (Neidle et al. 1997; Petronio and Lillo-Martin 1997; Baker-Shenk 1983, 1986), and rhetorical questions (Hoza et al. 1997), as well as many other syntactic and adverbial constructs (Anderson and Reilly 1998; Shepard-Kegl, Neidle, and Kegl 1995; Reilly, McIntire, and Bellugi 1990, Wilbur and Schick 1987; Coulter 1978, 1983; Liddell 1978; Baker-Shenk and Padden 1978; Friedman 1974).

In addition to these grammatical facial expressions and the full range of emotional facial expressions, which Ekman and Friesen (1975, 1978) contend are universal, both spoken and signed languages use facial expressions such as quizzical, doubtful, and scornful, which can be categorized as nonemotional and nongrammatical (NENG). These NENG facial expressions are commonly used during social interaction, without carrying emotional or grammatical meaning. We include them here in order to study a class of facial expressions that exhibits neither the automatic qualities of emotional expressions (Whalen et al. 1998) nor the structured and grammar-specific characteristics of ASL syntax described earlier.

ASL is a language of dynamic visuo-spatial changes that are often difficult to describe but nonetheless essential for our understanding of the language (Emmorey 1995). Grossman (2001) and Grossman and Kegl (submitted) emphasize the need to use *dynamic* facial expressions (video clips), as opposed to the commonly used *static* images (photographs), in order to obtain a more realistic assessment of the way in which hearing and deaf people recognize and categorize facial expressions. However, only a few detailed analyses of the production of dynamic emotional and grammatical facial expressions are available in ASL.

Baker-Shenk (1983) and Bahan (1996) have dealt extensively with the development of dynamic ASL facial expressions and their link to the manual components of ASL sentences. They have detailed their development and noted their onset, apex (maximal expression), duration of apex, and offset. Baker-Shenk and Bahan observed these dynamic changes in numerous ASL sentences and looked for common denominators among samples of the same type of expression (e.g., wh-question, y/n question) to determine how specific expression types differ from each other. Baker-Shenk used Ekman and Friesen's facial action coding system (Ekman and Friesen 1975, 1978) to analyze ASL question faces. In this system, each muscle group of the face is assigned an action unit (AU) number, and the specific combination of AUs defines a given facial expression. Using this technique, Baker-Shenk produced detailed descriptions of several different types of ASL question faces.

This approach, however, encounters some difficulties in describing dynamic features or gestures such as head tilts. For example, when looking at y/n questions, Baker-Shenk found that six samples out of sixteen had a downward head tilt, nine a forward tilt, and three had both. Despite those variations, she chose to call the downward head tilt the defining head tilt for y/n question expressions. The dynamic movements for wh-questions were even more variable, and she declined to make a specific assignment of head movement for these.

In addition to the difficulty of capturing the directional head movement variation within a given expression type, Baker-Shenk also tried to define the onset and offset speeds of the AU features she described, hypothesizing that a difference exists between the onsets and offsets of emotional and grammatical expressions. She admits, however, that those onsets and offsets exhibit strong variations, making certain emotional expressions appear like their grammatical counterparts and vice versa. The variations in the dynamic aspects of the facial expressions Baker-Shenk describes exemplify the difficulty in capturing the character of a class of expressions based only on descriptions of individual samples. Even an analysis as detailed and encompassing as Baker-Shenk's is hampered by variations between sentences, which make it nearly impossible to describe a definitive inventory of the expression type as a whole.

The detailed qualitative analyses conducted by Baker-Shenk and Bahan did not make use of SignStream (Baker-Shenk's work predated it), which is a multilevel database program developed as a collaborative effort between Boston University, Rutgers, and Dartmouth College (Neidle et al. 1997, 1998, 1999) and is designed specifically to code signed languages from dynamic video clips. SignStream can be used to track dynamic ASL features and gestures, such as eye aperture or body tilt, by describing the onset, apex, and offset of a given movement on a frame-by-frame basis and marking it according to a time index tied to the video clip. For example, the eyebrow raise during a surprise facial expression can be described as "starting" (frame A through frame B, i.e., beginning to rise from neutral at A and reaching a steady state at B). For the duration of the steady state (or apex), from frame B to C, it can be labeled with a specific value (e.g., "raised," "raised high," or "lowered"), allowing for the differentiation of various degrees of intensity. The same technique also codes the return from the steady state to baseline (labeled "end") from frame C to frame D. All of these events contain a minimum of two video frames but usually span a larger number of continuous frames. All of the dynamic features of ASL, including facial expressions, can thus be coded to capture both the physical and temporal dimension (figure 1).

Using this method, we can record data for the head movement, eyebrow movement, eye aperture changes, mouth movement, and

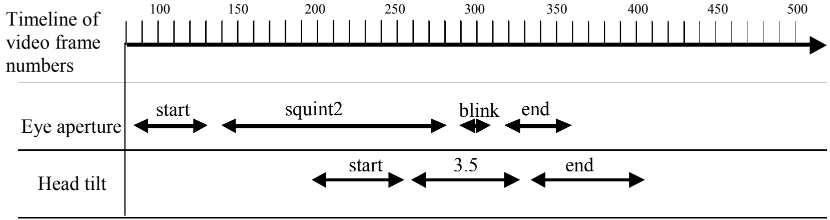



Figure 1. Schematic of SignStream dialogue window with coding for facial features.

so on for each expression. Typically these coded video clips are then examined for differences or similarities between expressions and their expression types. Although it is an excellent tool for coding complex aspects of ASL, SignStream does not yet enable us to go beyond a qualitative, narrative analysis of the features involved in producing individual grammatical, emotional, or NENG facial expressions.

The purpose of this article is to present a novel approach to the analysis of ASL facial expressions, using SignStream, to expand on its current capabilities, and to establish *quantitative* as well as qualitative analyses of facial expressions, thus bringing the study of facial expressions for the first time to a level where group comparisons, statistical analyses, and objective comparisons among facial expression types are possible.

## Materials and Method

### *Stimuli*

For this study we used six types of facial expressions as stimuli: neutral, angry, surprise, quizzical, y/n question, and wh-question. Neutral, angry, and surprise refer to the signing of a standard sentence with the different appropriate emotional overlays. The other stimulus types consist of a standard sentence being modified slightly to include either the phrase "I wonder" (quizzical) or a question marker (y/n question, wh-question). These stimulus types were chosen because they contain feature similarities but belong to different expression categories. Table 1 summarizes the similarities and category designation of the stimulus types.

Table 1. Stimulus Categories and Corresponding Facial Features

| | Stimulus Categories | | |
|---|---|---|---|
| Facial Features | NENG | Emotion | Grammar |
| Eye squint + furrowed brows | quizzical | angry | wh-question |
| Eyes widened + raised brows | | surprise | y/n question |
| Control condition | neutral | | |

*Stimulus Recording*

Stimuli were produced by two different signers (a woman, PLT, and a man, JCG) in separate recording sessions. Both signers participated in the creation and selection of their list of 20 stimulus sentences and ensured that the sentences were natural, colloquial ASL and comfortable to produce. The list of sentences varied between signers with some overlap (tables 2 and 3). After the complete set of stimuli was transcribed on paper, we taped JCG's and PLT's renditions of the six versions (neutral, angry, surprise, quizzical, y/n question, and wh-question) of each of the 20 sentences, resulting in 120 separately signed sentences each, for a total of 240 sentences in the database. Due to a recording error, one of PLT's sentences was lost, leaving 239 stimuli for study.

The stimuli were recorded with the camera focused on the signers' faces, capturing their facial expressions exclusive of hand and arm movements. The shots were monitored closely to ensure that no excessive mouthing of words occurred. A crucial aspect of the filming was to make certain that the signers' hands did not intrude into the camera's visual field since manual information would have corrupted the stimuli by supplying semantic content. The main difficulty in physically separating hand movements from facial actions is that many ASL signs are naturally signed with the hands near or on the face. For example, the word *who* is normally signed with the thumb of one hand placed on the chin. Since it would be impossible to establish a natural sample of ASL sentences excluding all signs produced at or near the face, the solution was to instruct JCG and PLT to "sign relaxed," a style of signing often used in informal social situations. During relaxed signing, the word *who* can be signed in front of the chest instead of at the chin, while maintaining the salient features of affective, grammatical, and NENG facial expressions.

With additional cameras we also captured the signers' torsos exclusive of the face, as well as the entire body. The torso and whole-body videos enabled us to corroborate the validity of each facial expression based on the corresponding manual component of the sentence. Since each of the 239 separate stimulus sentences was signed multiple times, we selected the best iteration of each to

TABLE 2. PLT's Stimulus Sentences in ASL Gloss and English

| | ASL Transcription | English Translation |
|---|---|---|
| 1 | MUST MEET PRINCIPAL AFTER SCHOOL FINISH | I have to meet the principal after school. |
| 2 | IX BOX MUCH HEAVY CARRY | This box is too heavy to carry. |
| 3 | HIS NEW CAR $20,000 | His new car cost $20,000. |
| 4 | MY CHILDREN THUP POPCORN | My children finished all the popcorn. |
| 5 | HER PARTY SAME-TIME MINE | Her party is the same time as mine. |
| 6 | IX TEACHER POSS FINAL EXAM TOUGH | This teacher's final exam is tough. |
| 7 | OWNER REFUSE SELL HOUSE | The owner refuses to sell the house. |
| 8 | MY NEW BYCYCLE BROKE | My new bicycle broke. |
| 9 | IX BOWL TABLE FALL-OFF | This bowl fell off the table. |
| 10 | MY FRIEND POSS DOCTOR MOVE-TO ENGLAND | My friend's doctor moved to England. |
| 11 | MY FAMILY VISIT THIS WEEKEND WILL | My family will visit this weekend. |
| 12 | MY FRIEND BORROW MY BRACELET | My friend borrowed my bracelet. |
| 13 | ALL MY CLASSES START EARLY | All of my classes start early. |
| 14 | MY CHILDREN BUY MUCH SODA | My children buy a lot of soda. |
| 15 | HE LOSE MY HISTORY BOOK | He lost my history book. |
| 16 | HE PAY $500 THAT DRESS IX | He paid $500 for that dress. |
| 17 | MY CAR RUN-OUT GAS | My car ran out of gas. |
| 18 | THAT MOVIE MUCH VIOLENT | That movie is very violent. |
| 19 | MY NEWSPAPER ARRIVE LATE | My newspaper arrived late. |
| 20 | I LEAVE KEYS IN CAR | I left the keys in the car. |

Notes: Wh-questions contained a wh-word at both the beginning and the end of the sentence.
Quizzical sentences contained ME WONDER at the beginning and the end of the sentence.
Y/N questions contained a manual question marker at the end of the sentence.

Table 3. JCG's Stimulus Sentences in ASL Gloss and English

| ASL Transcription | | English Translation |
|---|---|---|
| 1 | LATE ARRIVE NEWSPAPER | My newspaper arrived late. |
| 2 | FINISH WASH DISHES LAST NIGHT | I finished washing the dishes last night. |
| 3 | BUY BOOK THREE ME | I bought three books. |
| 4 | FINISH PAINT HOUSE LAST WEEK | I finished painting the house last week. |
| 5 | LOST MY BOOK | I lost my book. |
| 6 | FINISH READ BOOK HISTORY | I finished reading a history book. |
| 7 | VISIT MY FAMILY WEEKEND | I visit my family on weekends. |
| 8 | RUN-OUT GAS MY CAR | My car ran out of gas. |
| 9 | FINISH WALK FIVE MILES | I just walked five miles. |
| 10 | BROKE MY BICYCLE | I broke my bicycle. |
| 11 | CLEAN BATHROOM ONE HOUR | I cleaned the bathroom for one hour. |
| 12 | FINISH HOMEWORK | I finished my homework. |
| 13 | READY LEAVE NOW | I'm ready to leave now. |
| 14 | MUST BUY COMPUTER | I have to buy a computer. |
| 15 | HELP FRIEND MOVE | I'm helping a friend move. |
| 16 | WATCH TV ALL-DAY | I watch TV all day. |
| 17 | START EARLY MY CLASS | I start my class early. |
| 18 | BUY TOO-MUCH BANANA | I bought too many bananas. |
| 19 | SWIM ALL-DAY | I swim all day. |
| 20 | LEAVE KEY MY CAR | I left my keys in my car. |

Notes: Wh-questions contained a wh-word at the end of the sentence.
Quizzical sentences contained ME WONDER at the end.

include in the final database. This selection was based on the whole-body tape, which permitted us to evaluate the facial expression in connection with the underlying manual sentence. Each clip in the database was chosen only if both components, manual and facial, were produced correctly and without hesitation, were in focus at all times, and had no hand movements intruding into the facial space. Both signers had an opportunity to go over their productions on videotape and approve the stimuli. Once a whole-body clip was selected, we extracted the corresponding face-only clip and focused on the analysis of the facial expression independent of the underlying manual component. This allowed us to establish an unbiased inventory of onset, offset, and peak occurrences for each facial feature, which we could then process for group comparisons. Our aim was not to connect each facial movement to the position of its corresponding manual marker in the hierarchical tree but rather to establish a continuous timeline of facial feature movements for each expression type and to create a simple representation of those movements for group comparison of onset, offset, and peak occurrences.

### *SignStream Coding*

The stimuli were then coded in random order, and the coder was blind to the expression type label. In order to create a database in which the expressions were characterized consistently (and were thus comparable), we established specific guidelines to define the onset and offset of each event and to determine how to label each event apex or steady state.

### *Facial Features*

The ''start'' of every event was defined as lasting from the first frame of movement in the baseline until a steady state was reached. The ''end'' was defined as the sequence of frames that captured the return from the steady state to the baseline of the facial features (brow, cheek, etc.) in question. The intervening steady state was then given a specific label, such as ''squint,'' ''start,'' or ''brow raise'' and a corresponding numerical value for later analysis. The numerical values were chosen to reflect increases or decreases in intensity for each event as compared to their respective baseline. For example, ''eyes wide'' had a higher value than the eye aperture baseline, while ''squint'' was given a lower-than-baseline value. Table 4 shows all of the facial feature events defined in this database and their SignStream labels, as well as corresponding numerical values. Full head movements such as head tilt are not included in this table but are discussed later.

### *Cheeks*

The only type of cheek movement recorded was ''tensed.'' This refers to the cheeks moving up, without affecting the eye aperture.

Table 4. Inventory of Facial Expression, Facial Features, and Corresponding Labels

| Facial Features | | | | | | | |
|---|---|---|---|---|---|---|---|
| SignStream label for feature events | Eye aperture | Eyebrows | Cheeks | Mouth | Head shake | Head nod | Eye gaze |
| | Wide 8 | Raised 2 12 | Tensed 2 6 | Smile 6.5 | Single 1 | Up/down 1 | Right 0.5 |
| | Neutral 7 | Raised 11 | Tensed 5 | Smile only left 6 | Double 2 | Down/up −1 | Right/up 1 |
| | Squint 6 | Neutral 10 | Neutral 4 | Open 5.5 | | | Up 1.5 |
| | Squint 2 5 | Lowered 9 | | Frown 5 | | | Left/up 2 |
| | Squint 3 4 | | | Intense 4.5 | | | Left −0.5 |
| | Upper lid down 3 | | | Neutral 4 | | | Left/down −1 |
| | Blink1 | | | Raised upper lip 3.5 | | | Down −1.5 |
| | | | | Right lip up 3 | | | Down/right −2 |
| | | | | Pursed lips mm 2.5 | | | Off 0.2 |
| | | | | Pursed lips Mm-tight 2 | | | |
| | | | | Pursed lips Oo 1.5 | | | |
| | | | | Pursed lips Oo-tight 1 | | | |

While there are other possible cheek movements, such as "puffed," none were observed in the samples analyzed here.

### *Eye Aperture*

The different levels of "squint" ("squint," "squint2," and "squint3") were identified by how much of the iris was visible during the squint. Blinks were included in the SignStream coding but not in the final analysis. This exception was made because blinks are very brief events that are not connected to specific expression types but rather occur equally in all expressions.

"Upper lid down" was correlated mostly to changes in eye gaze. Eye gaze in ASL is often used in object-verb agreement and is therefore an integral part of a signer's facial actions but apparently is independent of the type of expression, whether grammatical, emotional, or NENG. Since a downward shift in eye gaze often causes the upper lids to droop, we found it necessary to establish a correlation between those two events in the data, prior to classifying "upper lid down" as a general feature unconnected to specific facial expression types and eliminating it from the final analysis.

JCG maintained virtually constant eye contact with the camera and had only very infrequent episodes of upper lid down. PLT, however, exhibited frequent eye gaze shifts in all expression types. Looking at a sample graph of eye gaze and eye aperture in PLT (figure 2), we clearly see that the sudden dips in eye gaze co-occur with sudden dips in eye aperture. We thus concluded that lowered eye gaze in PLT is strongly correlated with a decrease in eye aperture and independent of the underlying facial expression. Therefore, upper lid down was not considered a predictive value for any facial expression type and was not included in the final analysis.

### *Head Tilt*

The events that involved full head movement, such as head tilt, were recorded separately, using a grid that was superimposed on the screen displaying the expression video clip. In order to quantify head movement, we marked a zero point vertically along the axis of the nose and horizontally at the tip of the chin. Each video clip was situated so that the zero marks corresponded precisely to the nose and chin

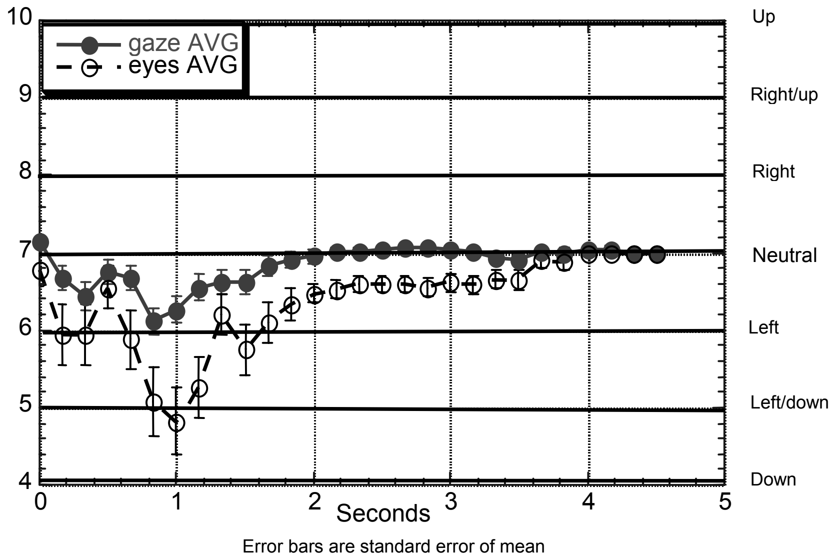


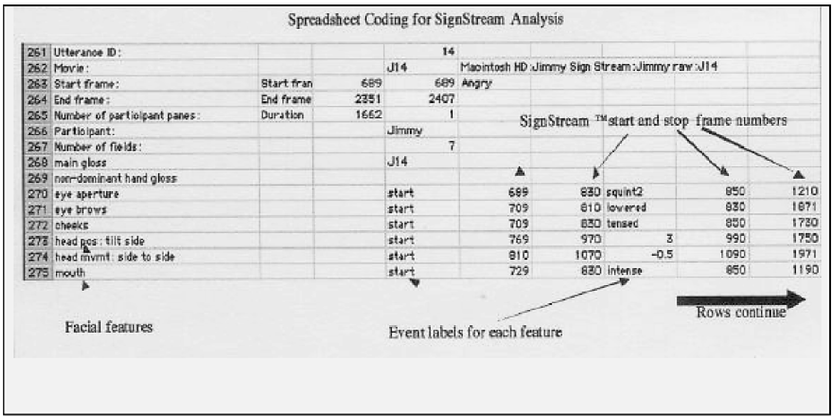

Figure 2. Eye aperture correlated with eye gaze in PLT's wh-question expressions.

lines of the face at rest, prior to the onset of the first facial feature of each expression. We were then able to quantify head movements by the number of coordinates a given part of the head deviated from its original position on the grid. For example, if the left temple was at vertical coordinate 2 before the expression started and then moved in a head shake to coordinate −2, the value for the steady state event of that movement was recorded as −4.[1] We chose different anchor points in the face for different types of head movements (e.g., the uppermost corner of the forehead was used for head tilts, the chin for head juts).

*Data Preprocessing*

Each event of the 239 samples was recorded, and the database was verified by a second coder to ensure accuracy and consistency. The raw SignStream data had to be modified significantly in order to create graphs that would represent quantifiable group data for facial features in the different expression types. This was accomplished by exporting the SignStream database to a spreadsheet program. In this new format, each facial expression stimulus was represented by a data cluster containing the start and stop frame number for the entire clip and a row of start and stop frame numbers for each event of a given feature in this sample (figure 3). For example, a surprise expression video clip would list a data row for eye aperture containing the beginning and end frame numbers for the start event, an eyes wide event, a blink, the continuation of eyes wide, and the end event, in sequence.

The next row in the cluster contained the same type of data for the eyebrow feature and so on. Each of the event labels was identified with its beginning and end frame numbers from the original movie file. Since the frame numbers are not absolute but rather are relative to the clip length and the placement of the actual expression within that clip based on the original SignStream analysis, they needed to be transformed for further analysis. In order to compare specific facial expression events across stimuli, it was necessary to choose a consistent and specific beginning frame for each expression. For that purpose, the first frame of the first recorded event in any expression was defined as the first frame of that expression. For example, if the facial expression's first event was an eye blink (which was frequently the case), then the first onset frame of that blink was also recorded as the onset frame of the entire expression. In the same manner, the final frame of each expression was determined to be the offset frame of the last recorded end event in each expression. Thus, the length of each expression was effectively determined as the time during which




| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 261 | Utterance ID: | | | 14 | | | | | |
| 262 | Movie: | | | J14 | Macintosh HD :Jimmy Sign Stream :Jimmy raw :J14 | | | | |
| 263 | Start frame: | Start fran | 689 | 689 | Angry | | | | |
| 264 | End frame: | End frame | 2351 | 2407 | | | | | |
| 265 | Number of participant panes: | Duration | 1662 | 1 | | | | | |
| 266 | Participant: | | | Jimmy | | | | | |
| 267 | Number of fields: | | | 7 | | | | | |
| 268 | main gloss | | | J14 | | | | | |
| 269 | non-dominant hand gloss | | | | | | | | |
| 270 | eye aperture | | | start | 689 | 830 | squint2 | 850 | 1210 |
| 271 | eye brows | | | start | 709 | 810 | lowered | 830 | 1871 |
| 272 | cheeks | | | start | 709 | 830 | tensed | 850 | 1730 |
| 273 | head pos: tilt side | | | start | 769 | 970 | 3 | 990 | 1750 |
| 274 | head mvmt: side to side | | | start | 810 | 1070 | -0.5 | 1090 | 1971 |
| 275 | mouth | | | start | 729 | 830 | intense | 850 | 1190 |

FIGURE 3. Spreadsheet coding for SignStream analysis.

any facial event occurred, and the frame numbers were recalculated to begin at zero for each stimulus.

Each data cluster containing information for specific feature movement was now identified with the expression type it represented, and we were able to group the clusters into the original six stimulus types. This created twelve separate databases, six each for JCG and PLT, with all of the normalized frame numbers and corresponding value labels for every facial feature event in every expression, grouped separately by expression type.

### *Establishing a Continuous Timeline*

To create a continuous timeline, we analyzed the spreadsheet using the numerical values for each facial feature event as a measure of intensity that could later be graphically rendered. At this stage of the analysis, there was a recorded sequence of events for each feature, such as the eye aperture feature, which was composed of events "start," "squint," "blink," and "end." However, the only frame numbers recorded by SignStream are the beginning and end frames for each feature event, so there was no complete, continuous time course for any sentence. In order to create a continuous timeline, we filled in the intervening time points with the same values recorded for their corresponding start and stop frames, using intervals of 100, in SignStream numbers (0.17 seconds) on a zero-to-five-second timeline, resulting in thirty data points. For example, if SignStream recorded an eyebrow lowered event beginning at frame 2,000 and ending at frame 2,300, the timeline was filled with the appropriate numerical value (9) for time points 2,000, 2,100, 2,200, and 2,300. If the start or stop frame number for an event was between two time points, such as 1,980 and 2,030, the number was rounded to the nearest 100. If no event was recorded for a specific time interval, the timeline was filled in with the neutral value for each feature event, such as 10 for neutral eyebrows, 7 for neutral eye aperture, or 0 for neutral head position. Using this method, we established complete time courses for each facial expression.

### *Averaging the Data*

Using these continuous timelines, we calculated the average event value for each of the thirty time points across all 20 samples in a given facial expression. This established a single, average time course of event value for each facial feature of a given stimulus type. These averages were calculated for all of the features that showed activity in at least 6 out of 20 samples. Movement of a specific facial feature in only 5 or fewer of the 20 samples in an expression type was deemed not representative of the expression type and was therefore not included in the final analysis. The averaged time courses were then transformed to represent seconds rather than SignStream frame numbers, and the standard error of mean was calculated for each average data point. The complete average time courses were then plotted to provide a graphic representation of all of the features involved in each facial expression type.

## Results

The results and discussion of the analysis are presented in individual subsections for each facial feature. Some features, such as eye aperture, were present in all or most of the expression types, while other features, such as head shake, appeared consistently in only one type of expression. We discuss the more prevalent features first and supplement those data with evidence from features that occur less frequently.

### *Sentence Length*

The average length of each stimulus was computed for both signers and ranged from 2.5 seconds to 4 seconds depending on signer and sentence type. The difference in average sentence length between signers within a given sentence type ranged between 0.0 seconds and 1.0 second.

Since it is impossible to produce 239 natural sentences of exactly the same length, a temporally homogeneous sample could not be collected. It would have been possible to create a more homogenous sample by artificially compressing and expanding the digitized sentences to a specific, median target time. However, such a modification would have distorted the crucial timing information contained in the natural production of the sentences. The analysis presented here highlights several small but important timing differences that distinguish expression types. These differences would have been lost if we had artificially adjusted the sentence length of the samples. The data also show that most of the differences occur at the beginning of the sentence, making it less relevant whether the endpoints of the sentences match up.

*Eye Aperture and Eyebrow Movement*

The data show that there are virtually no changes in either eye aperture or eyebrow movement during a neutral, nonemphasized sentence, indicating that feature movement does not occur randomly in nongrammatical, nonemotional, nonemphasized sentences. This holds for the other facial features we analyzed as well. Looking at all of the other expression types we find that eye aperture and brow level are modulated for each expression type and also are always modulated in synchrony. There are no expressions in which the eyebrows are raised while the eye aperture is decreased or vice versa (figures 4a and 4b).

JCG tends to have faster and more pronounced onsets of both features for some expression types than PLT, although those differences may simply reflect the fact that JCG's sentences were somewhat shorter overall than PLT's. It is important to note that

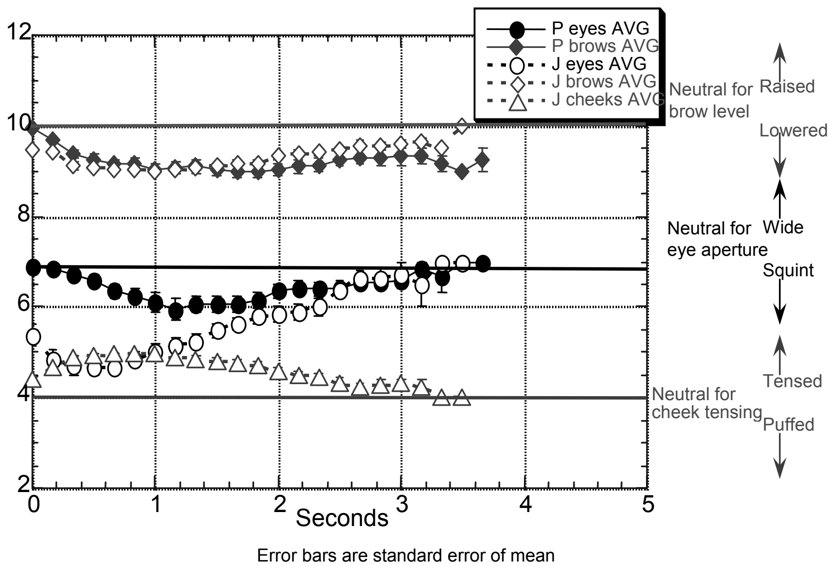



Figure 4a. Eye aperture, brow level, and cheek movement for PLT and JCG in angry expressions.

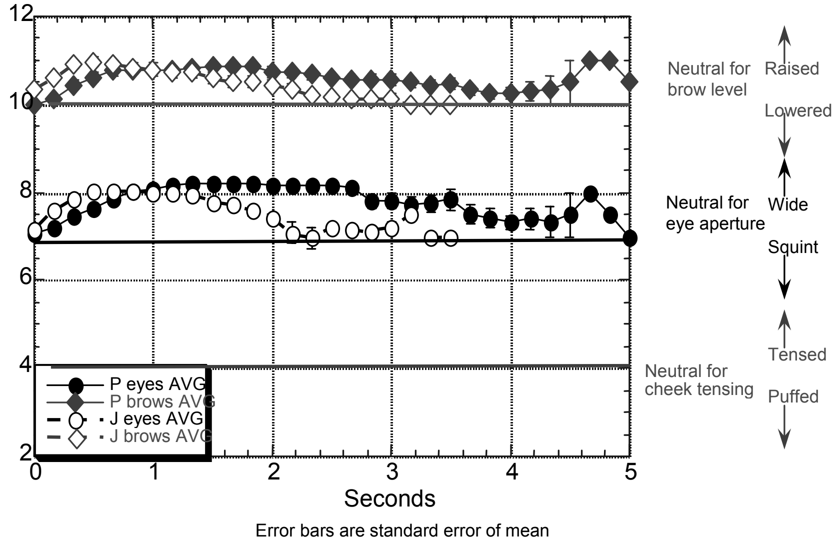



FIGURE 4B. Eye aperture, brow level, and cheek movement for PLT and JCG in surprise expressions.

differences in the degree of eye aperture were more easily monitored and coded than those in degree of eyebrow movement. It was nearly impossible to clearly define varying degrees of brow level changes, mostly because of a lack of physical markers around the brows against which to measure movement. In contrast, eye aperture was easily differentiated by the amount of iris or pupil that was visible between the lids. As a result, there were fewer labels for brow movement than for eye aperture, which explains why there are more intensity variations recorded for eye aperture than for brow movements.

Overall, angry, wh-question, and quizzical expressions exhibit lowered brows and squinted eyes, while surprise and y/n question expressions show raised brows and widened eyes. The eye aperture graph for angry, quizzical, and wh-question expressions also shows that JCG has faster and stronger eye aperture decreases (squints) than PLT for all three expression types (figure 5a).

Beyond this simple description, the more interesting question is whether we can distinguish between expressions that appear similar, based on their specific eye aperture and brow level. The data also show that angry expressions exhibit faster and stronger squints than

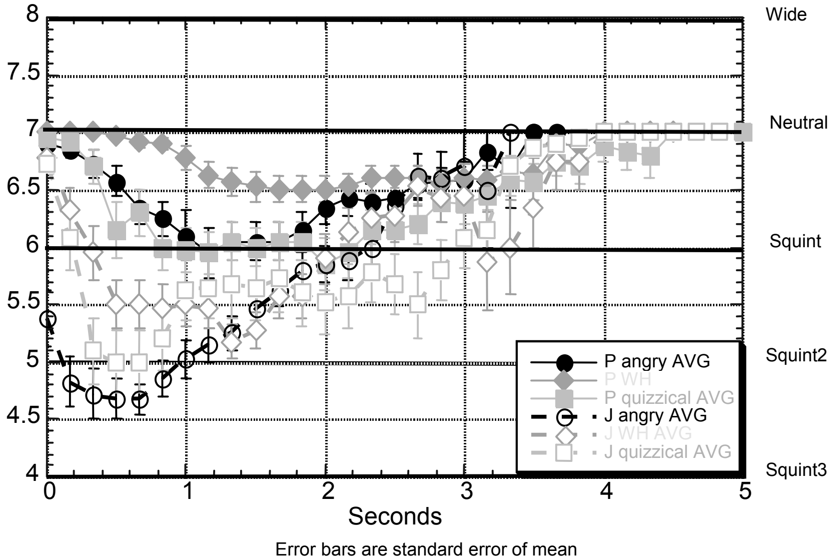



Figure 5a. Eye aperture for angry, quizzical, and wh-question faces: PLT and JCG.

wh-question faces in both PLT and JCG. So, even though JCG's eyes tend to show more rapid and pronounced squints than PLT's in all expressions, the relationship between angry and wh-question eye aperture events is maintained within each signer. In other words, the angry expressions in both signers show faster onsets and higher intensity for eye aperture and faster onsets for brow movements than their respective wh-question faces.

A clear rise-and-fall pattern occurs in the eye events in angry expressions and wh-questions in both signers, especially JCG, who exhibits a sharp rise to an apex in eye aperture and brow movements during the first second of the expression. Both signers rise to an apex in their brow events for all three expression types in the first two seconds of the sentence (figure 5b).

Quizzical expressions, on the other hand, are more prolonged and, in the case of JCG, exhibit a characteristic double apex pattern found only in these expressions. PLT's brow data show a flatter distribution (they stretch over the entire sentence), which is clearly different from the more rapid rise and fall in the other expression types.

In summary, eye aperture and eyebrow levels always move in synchrony, and angry expressions show faster and often more extreme

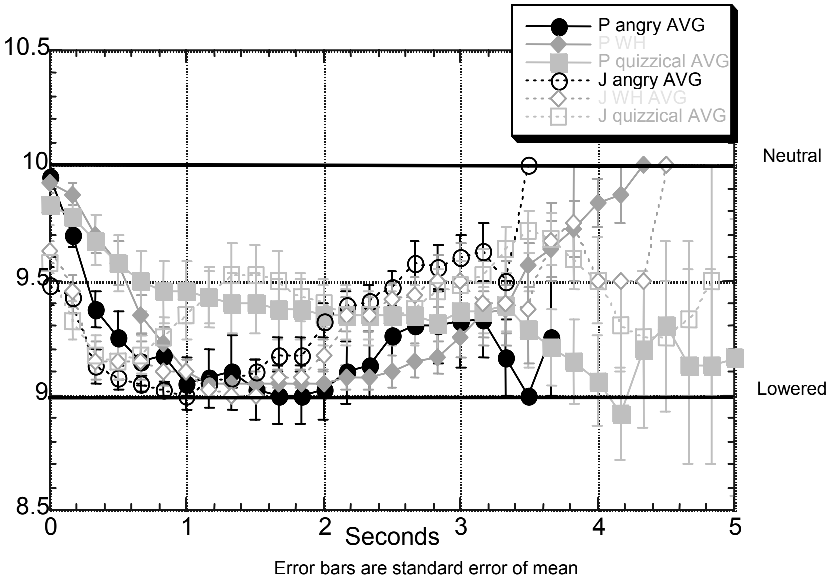



FIGURE 5B. Brow level for angry, quizzical, and wh-question faces: PLT and JCG.

changes in these features than the superficially similar wh-question faces. Quizzical expressions show a slower, longer, flatter, and even a double apex pattern of changes in aperture and brow level than their emotional or grammatical counterparts. These differences among the three superficially similar expression types were observable in both signers.

The brow raise graphs for y/n question faces and surprise expressions show the same internal relationships between grammatical and emotional faces as the brow raises in the previously observed triad of expressions (figure 6a). Although JCG's brows move faster than PLT's for both stimulus types, both PLT and JCG show an earlier apex and a faster return to baseline during the emotional expression (surprise) than in the grammatical (y/n question).

The eye aperture graph for these two expression types shows a pattern that is slightly less clear (figure 6b). JCG's eyes become wider more quickly during the surprise samples than during the y/n question samples, but during both expression types the eye aperture relaxes to baseline at the same time. PLT's eye aperture shows not only a faster onset and more extreme widening for the surprise expressions but also a slower decline than during the y/n question samples.

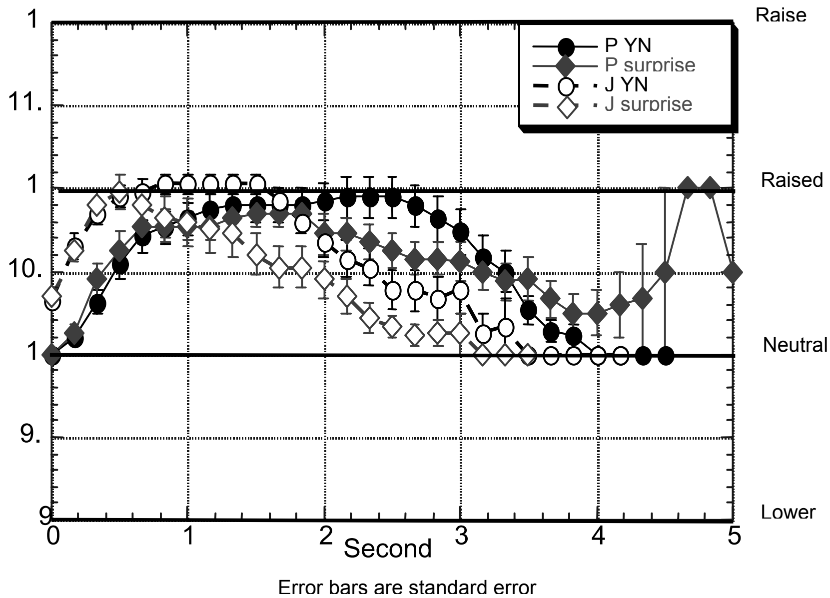



Figure 6a. Brow raise in surprise and y/n question faces: PLT and JCG.

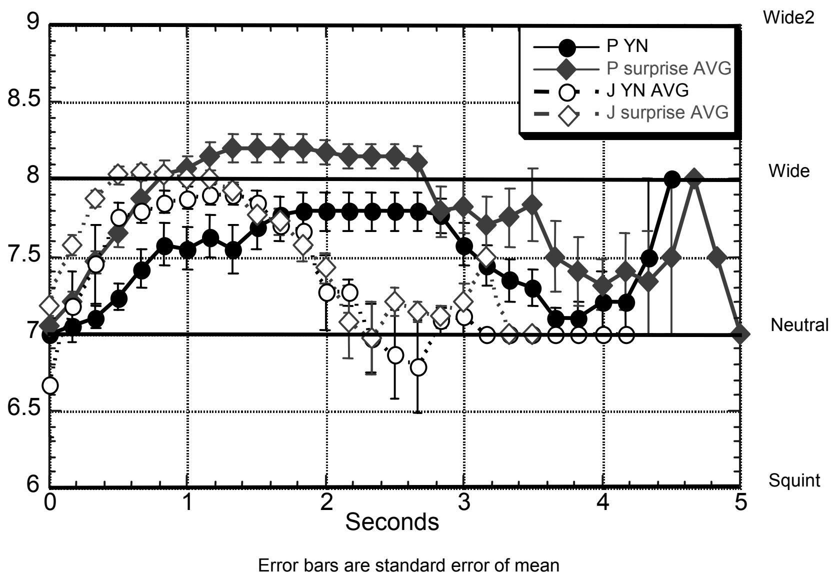



Figure 6b. Eye aperture in surprise and y/n question faces: PLT and JCG.

In summary, surprise shows faster and, in the case of eye aperture, more extreme movement than y/n questions, while there seems to be less regularity in the length of time that these features remain at their maximum level or the relaxation to baseline.

### *Cheeks*

The most obvious finding for the cheeks is that PLT virtually never uses them, while JCG employs them consistently for angry, quizzical, and wh-question expressions (figure 7). While this is a significant difference between PLT and JCG, it is also further evidence that these three expression types share feature similarities within a given signer.

JCG's cheek tensing exhibits the same relationship among emotional, grammatical, and NENG stimulus types as the eyebrow and eye aperture movements. During angry expressions JCG's cheeks tense sooner and more briefly than during wh-questions, reaching a pronounced apex during the first second and then falling more rapidly. On the other hand, wh-questions reach their cheek-tensing apex only during the subsequent second and fall more gradually. The

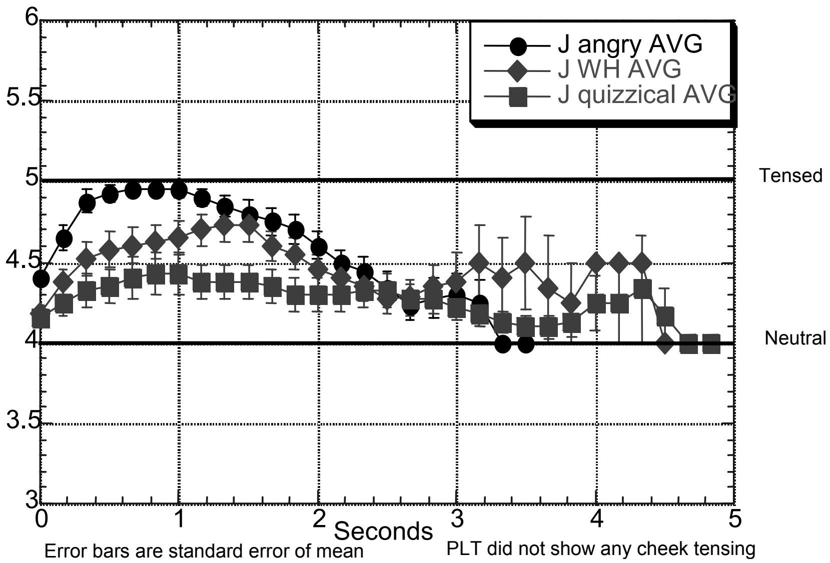



FIGURE 7. Cheek tensing for angry, quizzical, and wh-question faces: JCG.

quizzical expressions show the typically less intense and more sustained, flatter pattern over the course of the sentence.

*Mouth*

The mouth reflects the greatest number of differences between the two signers in our database. The raw video clips reveal that PLT is more apt to mouth words than JCG, marking a significant difference in signing styles. In the graphs we compiled, PLT's mouthing of words was coded as "neutral," as long as it was not combined with another mouth event, such as "lips pursed" or "frown." This placed JCG and PLT on somewhat more equal footing for the sake of comparison, but many individual differences remain. The data for mouth movement in all of the expression types for JCG and PLT show no discernible patterns or similarities in signers or across expression types. The only exception occurs in the quizzical samples (figure 8), in which both signers show lip pursing in the characteristic double apex pattern, with two clearly separate apices toward the beginning and end of the sentence.

JCG shows more pronounced or consistent lip pursing throughout the sample sentence than PLT, but the pattern and type of movement are quite similar. Although overall mouth movements for all

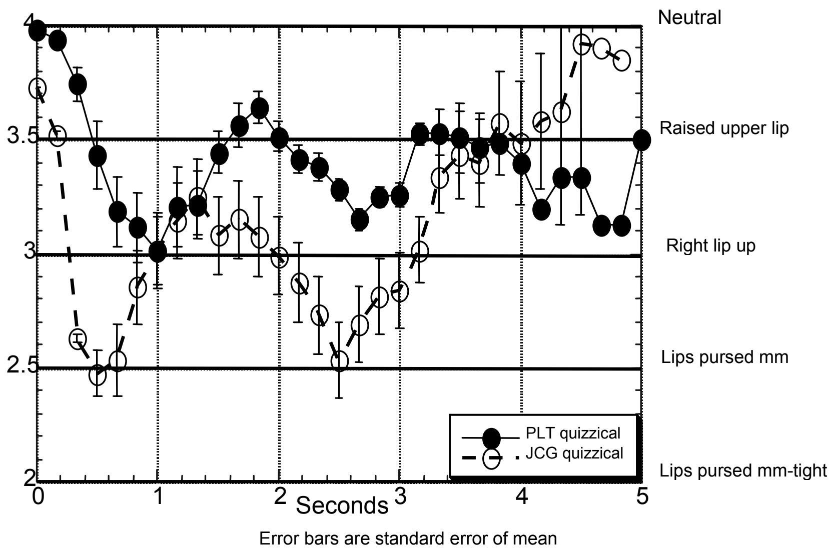



Figure 8. Mouth movement for quizzical expressions: PLT and JCG.

of the other expression types are highly variable, they are strongly predictive for quizzical expressions, where the characteristic lip pursing occurs consistently in the double apex pattern observed in other facial features of this stimulus type.

*Head Tilt, Jut, and Side-to-Side Movement*

Side-to-side movement and head jut do not yield a discernible pattern for any expression type. Head tilt, however, seems to be a more distinguishing feature for wh-question, angry, and quizzical expressions (figure 9). While wh-questions show only a slight degree of head tilt in a flat distribution across the sentences, angry expressions in both signers include a sharp rise to and fall from a head tilt apex within the first second, with a subsequent flatter and less intense distribution over the rest of the sentence. This early spike is consistent with the behavior of other features during emotional expression types.

The quizzical expressions show the most pronounced head tilt in both signers, who chose different tilt directions but a similar movement pattern. Both have a smooth distribution of intensity across the sentences at a significantly higher level of tilt than the grammatical

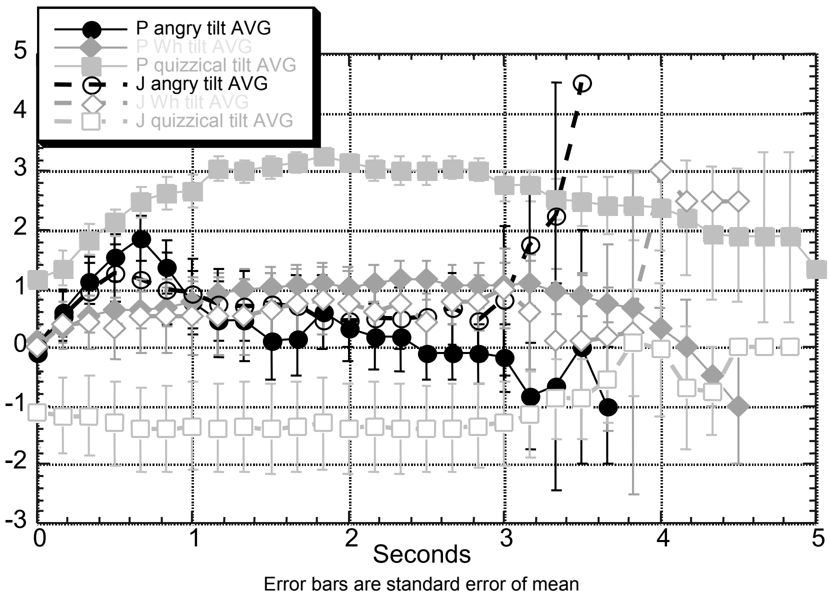



FIGURE 9. Head tilt in angry, quizzical, and wh-question faces: PLT and JCG.

and emotional samples in this grouping, indicating a moderate head tilt sustained over the whole sentence.

In summary, the only predictive patterns in this group of features are found in the head tilt graph for angry, quizzical, and wh-question faces. In this comparison, head tilt shows a rapid, early, and sharp rise to and fall from the apex for the emotional expression, consistent with data from other facial features and in contrast to the flat distribution of head tilt found in the grammatical counterpart. The quizzical expressions are characterized by a significantly more pronounced head tilt that extends over the entire sentence.

*Shoulders*

Shoulder raises are not observed in most expression types. The three types (angry, surprise, and wh-question) that do contain them show much larger standard errors of mean than other feature types, indicating that shoulder raise is not a consistent, or even necessary, component of any expression type.

*Head Shake and Nod*

The head nod data for quizzical expressions do not deviate substantially from zero, but it is important to note that quizzical sentences were the only ones that contained a consistent head nod in both signers. Neither signer used a consistent head nod in any other type of expression. The fact that both signers' nods are evenly distributed over almost the entire sentence is consistent with the pattern of feature distributions in quizzical expressions in general.

The head shake data for wh-question faces show a somewhat more significant deviation from zero, especially for PLT. Again, the most important finding for this feature is that wh-questions are the only expression types to contain a consistent head shake at all. Both subjects show a head shake distribution over almost the entire question, with two clearly definable apices that are close together and centered over the middle of the sentence (figure 10).

When we include these data in the comparison of angry, quizzical, and wh-question faces, it becomes clear that head nod and head shake represent two of the most salient features in distinguishing among the three stimulus types. Although all three expression categories contain

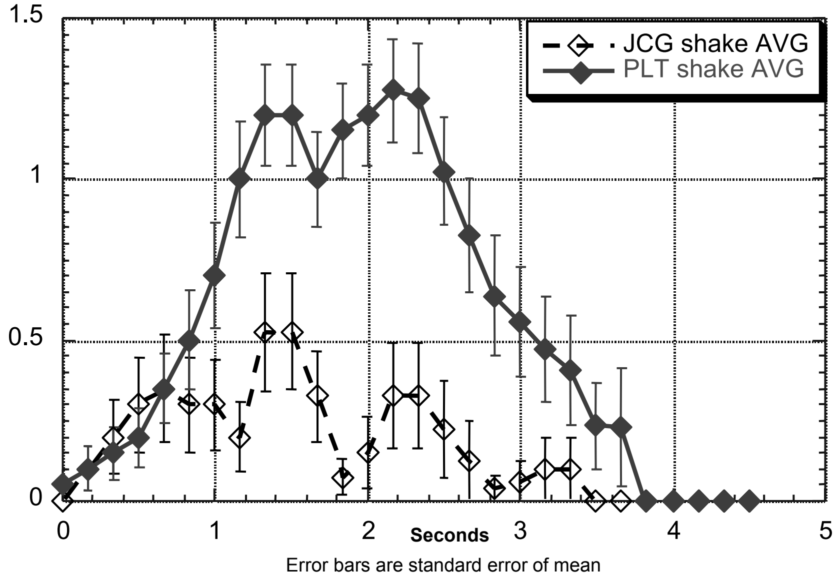



FIGURE 10. Head shake in wh-question faces: PLT and JCG.

grossly similar eye aperture, brow, and, in one signer, cheek movements, quizzical is the only one of the three expression types with a consistent head nod; moreover, the wh-question is the only expression with a consistent head shake. Although these differences are easily discernible in a video clip analysis of facial expressions, they are impossible to portray in an analysis using still photographs.

The consistency of the head shakes in wh-questions and the predictability of onset, offset, and peak head shake occurrences across both signers add important information to our understanding of the grammatical nature of this facial expression. McClave (2001) notes the existence of a wh-question head shake in ASL but refers to it as a conventionalized head gesture borrowed from the hearing population. She claims that head shake has not yet reached the level of grammaticization since it does not exhibit a predictable onset and offset pattern. The data shown in figure 10 appear to contradict that statement, based on the small sample of signers we analyzed. The onset and offset patterns of the wh-question faces show predictability between signers, extending even to the timing of the double peaks in the center of the sentences. This regularity seems to indicate that head shakes are an integral and necessary part of the grammatical structure of ASL wh-questions.

*Peak Realization*

Grossman (2001) emphasizes the importance of researching dynamic stimuli rather than static images of ASL facial expressions, and our analysis supports that statement. The head shake and head nod features that define the wh-question and quizzical stimuli cannot be seen in a static representation of those expressions. For individual features, the differences in onsets between emotional expressions and their grammatical counterparts are preserved only in dynamic presentations. Eliminating this timing information (by using static images) may significantly alter a subject's ability to distinguish among different types of expressions. Furthermore, and maybe most importantly, the graphs presented here clearly show that features making up a given facial expression have their peak realizations at different points in time. In view of the fact that each data point on our graphs corresponds to 5.7 frames of the original movie file (based on a 0.17-second time interval and video resolution of 30 frames per second), it is quite clear that no single video frame captures the peak realizations of all of the features.

*Emphatic Stress*

The dynamic changes that occur in ASL manual signs when produced with emphatic stress were originally documented by Friedman (1974) and elaborated upon by Wilbur and Schick (1994) and Coulter (1990). Emotions such as anger and surprise were used to create emphatic stress in an utterance (Wilbur and Schick 1987). Research has shown that the timing and spatial expanse have a quantifiable effect on manual signs under those conditions (Reilly, McIntire, Seago 1992). When stressed, signs tend to become faster and often shorter, with variation in hold and other spatial patterns. All of these studies focused on movement and timing changes of manual signs and paid only peripheral attention to the fact that emphatic stress causes more forceful facial expressions (Wilbur and Schick 1987).

In order to show that facial expressions follow a similar pattern, we measured the time points of the peak realizations for every facial feature in each expression type. While there were significant variations between the two signers in absolute timing of peaks, the relationship between the peaks of features during emotional stimuli and those occurring during their related grammatical counterparts were quite similar in both PLT and JCG. Specifically, we contrasted the peaks of features during angry expressions with those produced during wh-questions. The same comparison was done between the surprise and the y/n question faces. On average, the peaks for features in JCG's two emotional stimulus types occurred 0.56 seconds faster than in their grammatical counterparts. For PLT, the difference was 0.5 seconds. Although this evidence is still preliminary, it indicates that the faster onsets and generally more clipped transitions of manual signs produced under emphatic stress are carried over into their corresponding facial expressions.

## Summary

The data in table 5 illustrate the group differences between expression types (e.g., y/n question, surprise, wh-question, angry, quizzical) that share many features. These data represent a summary of facial expressions based on the average feature movements of the different expression types, rather than descriptions of isolated expression samples.

Surprise and y/n question expressions share the same features but differ in the speed and length of the rise to and fall from the apex for eye aperture and eyebrows. The emotional expression shows a faster, more pronounced rise and fall pattern than its grammatical counterpart.

Angry, quizzical, and wh-question expressions share many—but not all—of the features. Again, the emotional expression shows a faster and more pronounced rise to and fall from the apex for eye aperture and eyebrow than the grammatical expression in this triad. The same holds true for cheek tensing for the signer who used it (JCG). Quizzical expressions show a sustained flat distribution or double apex for those same features. Moreover, quizzical and angry expressions have a consistent head tilt (the angry tilt shows a sharp, early rise to and fall from an apex, whereas the quizzical tilt has a

Table 5. Summary of Salient Features for All Expression Types

| Stimulus Type _ | Eye Aperture | Eye Gaze | Brow Level | Cheek Mvmt. | Mouth Mvmt. | Head Tilt | Head Jut | Head Side-to-Side | Head Nod | Head Shake | Shoulder Mvmt. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| neutral | neutral | ⋆ | neutral | neutral | ⋆ | ⋆ | ⋆ | ⋆ | — | — | — |
| angry | fast, early, sharp apex squint | ⋆ | fast, early, sharp apex lowered | *fast, early, sharp apex tensing* | ⋆ | fast, early, sharp apex | ⋆ | ⋆ | — | — | ⋆ |
| surprise | fast, early, sharp apex wide | ⋆ | fast, early sharp apex raised | — | ⋆ | ⋆ | ⋆ | ⋆ | — | — | ⋆ |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| quizzical | sustained flat or double apex squint | *broken for entire sentence* | sustained flat or double apex lowered | *sustained flat apex tensing* | sustained double apex lips pursed | sustained flat apex strong tilt | ⋆ | ⋆ | sustained flat apex | — | — |
| y/n question | slower, medium apex wide | ⋆ | slower, medium apex raised | — | ⋆ | ⋆ | ⋆ | ⋆ | — | — | — |
| wh-question | slower, medium apex squint | ⋆ | slower, medium apex lowered | *slower, medium apex tensing* | ⋆ | ⋆ | ⋆ | ⋆ | — | centered, sharp apex | ⋆ |

Notes: Values in italics are found in only one of the two signers.
— means the value was not found in the expression type.
⋆ means the value was not predictive of expression type.

more sustained rise and fall pattern in which the maximum feature excursion is maintained longer).

In addition to those quantitative differences, quizzical sentences are the only expressions in the sample with a consistent and sustained head nod as well as consistent lip pursing in the double apex pattern characteristic of this expression type. Moreover, PLT shows a consistent and sustained eye gaze break for these expressions. Wh-questions, on the other hand, are the only expressions in the sample with a consistent head shake, showing a clear, centered apex over the entire sentence.

## Discussion

The SignStream coding, subsequent spreadsheet analysis, and graphing of the six stimulus types discussed here enabled us to clearly depict the differences between emotional, grammatical, and NENG facial expressions. These data also provide a glimpse into the nature of differences between native signers' productions of these facial expression types. However, because we have data from only two signers, it is not possible for us to draw generalized conclusions. The results presented here indicate that certain features such as eye gaze and cheek and shoulder movements are not required elements of specific facial expressions; instead, they are used only by some signers. When these features are used, however, they seem to follow the same temporal patterns as other features in the same expression type.

The differences between emotional and grammatical expression types that share many features lie mainly in the speed and development of the rise and fall of the intensity of a particular feature. Specific expressions contain features that are not shared by any other expression type in this sample and thus constitute a salient distinguishing element of those expressions (e.g., the characteristic head shake in wh-question faces, the equally characteristic lip pursing and head nod in quizzical expressions). By looking at a simple graphic depiction of our data, we can now also predict the occurrence of selected facial movements in a specific expression type, thereby making claims about the grammatical necessity of facial gestures heretofore believed to be simply conventionalized but not yet grammaticized.

Our technique for coding facial expressions is a novel method that enables us to describe and quantify differences between similar-appearing expression types as expressed in group averages. Because of the complex, multilevel, and dynamic nature of facial expressions, it is necessary to use a coding system such as SignStream to capture all of the nuances inherent in the expressions. However, SignStream coding provides a great deal of data in a format that is not yet suitable for group comparisons. The side-by-side assessment of individual video clips for which SignStream is currently designed is not sufficient to achieve quantitative and accurate group characterizations of facial expression types. The technique introduced here allows the consolidation of data for all aspects of facial expressions—temporal as well as spatial, quantitative as well as qualitative—into a format that makes simple calculations and a graphic, time-correlated display of those data possible. The graphs and tables included here represent a clear depiction of these types of complex data in a summary format that is easily accessible for those who were not involved in the actual coding process and may not have seen the raw stimuli. This type of analysis can contribute greatly to our ability to make group comparisons between different facial expression types and understand how they are produced.

## Acknowledgments

The authors would like to thank Anthony Barrand, Wayne Cowart, Robert Savoy, and Romy Spitz. We thank the Deaf signers who produced the stimuli; Rebecca Damon, Nat Ives, Leila Monaghan, MaryAnn Nilsson, and Romy Spitz for camera and editing work; and the Media Studies Program of the University of Southern Maine for the use of their equipment. We also acknowledge NSF grant #SBR-929065 from the University of Southern Maine, "The Architecture of Functional Categories in American Sign Language," and NSF grant #SBR-9996330 from the University of Southern Maine, "The Emergence of a Signed Language in Nicaragua," for cross-linguistic research.

## Note

1. A head shake represents a continuous, rather rapid and often repeated back-and-forth turning of the head, while a side-to-side movement is defined as an initial turn followed by a sustained steady state in the turned position, followed by a return to baseline.

## Bibliography


Aarons, D. 1994. Aspects of the Syntax of American Sign Language. PhD diss., Boston University.

———, B. Bahan, J. Kegl, and C. Neidle. 1992. Clausal Structure and a Tier for Grammatical Marking in American Sign Language. *Nordic Journal of Linguistics* 15(2): 103–42.

Anderson, D. E., and J. S. Reilly. 1998. PAH! The Acquisition of Adverbials in ASL. *Sign Language and Linguistics* 1(2): 117–42.

Bahan, B. 1996. Nonmanual Realization of Agreement in American Sign Language. PhD diss., Boston University.

Baker-Shenk, C. 1983. A Microanalysis of the Nonmanual Component of Questions in American Sign Language. PhD diss., University of California–Berkeley.

———. 1985. The Facial Behavior of Deaf Signers: Evidence of a Complex Language. *American Annals of the Deaf* 130(4): 297–304.

———. 1986. Factors affecting the Form of Question Signals in American Sign Language. In *Diversity and Diachrony,* ed. D. Sankoff, 407–14. Philadelphia: John Benjamins.

———, and C. Padden. 1978. Focusing on the Nonmanual Components of ASL. In *Understanding Language through Sign Language Research: Perspectives in Neurolinguistics and Psycholinguistics,* ed. P. Siple, 27–57. New York: Academic Press.

Coulter, G. R. 1978. Raised Eyebrows and Wrinkled Noses: The Grammatical Function of Facial Expressions in Relative Clauses and Related Constructions in ASL in a Bilingual, Bicultural Context. Paper presented at the Second National Symposium on Sign Language Research and Teaching, Coronado, Calif., National Association of the Deaf.

———. 1983. A Conjoined Analysis of American Sign Language Relative Clauses. *Discourse Processes* 6(3): 305–18.

———. 1990. Emphatic Stress in ASL *Theoretical Issues in Sign Language Research,* ed. S. Fischer and P. Siple. Vol. 1: *Linguistics,* 109–26. Chicago: University of Chicago Press.

Ekman P., and W. V. Friesen. 1975. *Unmasking the Face: A Guide to Recognizing Emotions from Clues*. Englewood Cliffs, N.J.: Prentice-Hall.

———. 1978. *Facial Action Coding System*. Palo Alto, Calif.: Consulting Psychologists Press.

Emmorey, K. 1995. Processing the Dynamic Visual-spatial Morphology of Signed Languages. In *Morphological Aspects of Language Processing,* ed. L. B. Feldman, 29–54. Hillsdale, N.J.: Erlbaum.

Friedman, L. 1974. On the Physical Manifestation of Stress in the American Sign Language. Unpublished manuscript, University of California–Berkeley.

Grossman, R. 2001. Facial Expressions in American Sign Language: Behavioral, Neuro-imaging, and Facial Coding Analyses. PhD diss., Boston University.

———., and J. Kegl. Moving Faces: Categorization of Dynamic Facial Expressions in American Sign Language by Deaf and Hearing Subjects. *Journal of Nonverbal Behavior* (submitted).

Hoza, J., C. Neidle, D. MacLaughlin, J. A. Kegl, and B. Bahan. 1997. A Unified Syntactic Account of Rhetorical Questions in American Sign Language Syntactic Structure and Discourse Function: An Examination of Two Constructions in American Sign Language. American Sign Language Linguistic Research Project Report 4. Boston: Boston University.

Liddell, S. 1978. Nonmanual Signals and Relative Clauses in American Sign Language. In *Understanding Language through Sign Language Research*, ed. P. Siple, 55–90. New York: Academic Press.

———. 1986. Head Thrust in ASL Conditional Marking. *Sign Language Studies* 15(52): 243–62.

McClave, E. 2000. Linguistic Functions of Head Movements in the Context of Speech. *Journal of Pragmatics* 32(7): 855–78

———. 2001. The Relationship between Spontaneous Gestures of the Hearing and American Sign Language. *Gesture* 1(1): 51–72

Neidle, C., O. Foelsche, J. Kegl, D. MacLaughlin, B. Bahan, and D. Greenfield. 1997, 1998, 1999. SignStream Software v. 1.5. Trustees of Dartmouth College, Boston University, and Rutgers, the State University of New Jersey.

Neidle, C., J. A. Kegl, D. MacLaughlin, B. Bahan, and R. G. Lee. 2000. *The Syntax of American Sign Language: Functional Categories and Hierarchical Structure*. Cambridge, Mass.: MIT Press.

Shepard-Kegl, J., C. Neidle, and J. A. Kegl. 1995. Legal Ramifications of an Incorrect Analysis of Tense in ASL. *Journal of Interpretation,* Special Issue 7, *The Bilingual/Bimodal Courtroom:* 53–70.

Whalen, P., S. L. Rauch, N. L. Etcoff, S. C. McInerney, M. B. Lee, and M. A. Jenike. 1998. Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity without Explicit Knowledge. *Journal of Neuroscience* 18(1): 411–18.

Wilbur, R. B. 1994. Stress, Focus, and Extrametricality in ASL. Unpublished manuscript, Boston.

———, and B. Schick. 1987. The Effects of Linguistic Stress on ASL Signs. *Language and Speech* 30(4): 301–23.