J Nonverbal Behav
DOI 10.1007/s10919-006-0022-2

ORIGINAL PAPER

Moving Faces: Categorization of Dynamic Facial Expressions in American Sign Language by Deaf and Hearing Participants

Ruth B. Grossman · Judy Kegl

Abstract American Sign Language (ASL) uses the face to express grammar and inflection, in addition to emotion. Research in this area has mostly used photographic stimuli. The purpose of this paper is to present data on how deaf signers and hearing non-signers recognize and categorize a variety of communicative facial expressions in ASL using dynamic stimuli rather than static pictures. Stimuli included six expression types chosen because they share overt similarities but express different content. Hearing participants were more accurate in their categorizations but expressed overall lower confidence regarding their performance.

Keywords ASL · Dynamic facial expressions · Categorization · Accuracy · Confidence

Introduction

American Sign Language (ASL) requires the use of the face not only to express the full range of emotional facial expressions (Ekman & Friesen, 1975, 1978), but also to mark a large variety of language-specific grammatical constructs, such as topics (Aarons, 1996), agreement (Bahan, 1996; MacLaughlin, 1997; Neidle, MacLaughlin, Bahan, & Kegl, 1996), and several different kinds of questions: wh-questions (questions using who, what, where, when or why), yes/no questions (Baker-Shenk, 1983, 1986; Neidle, MacLaughlin, Bahan, Lee, & Kegl, 1997; Petronio & Lillo-Martin, 1997), and rhetorical questions (Hoza, Neidle, MacLaughlin, Kegl, & Bahan, 1997). Additionally, both spoken and signed languages use facial expressions, such as quizzical, doubtful, and scornful expressions, that accompany natural conversational

R. B. Grossman (✉)
Lab of Developmental Cognitive Neuroscience, Boston University School of Medicine,
715 Albany Street, L-814, Boston, MA 02118, USA
e-mail: ruthberg@bu.edu

J. Kegl
University of Southern Maine, Portland, ME, USA

interactions. When discussing these different classes of facial expressions, we mostly describe the feature involvement of each expression in its most pronounced static state. Several of these facial expressions have significant overlap in their feature involvement, indicating that, at least superficially, they look very much alike. For example, research has established that wh-question faces and angry expressions share many features, such as a furrowed brow and squinted eyes, while yes/no question faces and surprised expressions share raised eyebrows and widened eyes (Baker-Shenk, 1983, 1986). Data also exist in the literature to suggest that quizzical facial expressions share a significant number of features with wh-question and angry faces (Petronio & Lillo-Martin, 1997). These studies largely neglect the essential and natural dynamic nature of facial expressions.

ASL is a language of dynamic visuo-spatial changes that are often difficult to describe, but essential for our understanding of the language (Emmorey, 1995). The same holds true for facial expressions used during spoken English conversation. Ekman (1984) pointed out that the durations of various dynamic aspects of a facial expression (latency, apex, offset) provide valuable information about the intensity and potentially the type of facial expression displayed. These dynamic components can inform our recognition of facial expression type even when the facial features are obscured. In a point-light paradigm that masked most facial features, Bassili (1978, 1979) found that accuracy for categorization of different facial expression types was greater when the dynamic progression of the expression was preserved. These data emphasize the intuitive notion that the natural dynamic presentation of facial expressions has a profound effect on our ability to recognize and categorize such expressions in a research environment.

Only a few detailed analyses of natural productions of dynamic emotional and grammatical facial expressions in ASL have been conducted. Baker-Shenk (1983), used the Facial Action Coding System (FACS, Ekman & Friesen, 1975, 1978), and Bahan (1996) dealt extensively with the dynamic development of ASL facial expressions produced by deaf signers and their link to the underlying syntactic (manual) structure, detailing their development and noting their onset, apex, duration of apex, and offset. Their analyses focused exclusively on the production of ASL facial expressions, as opposed to recognition or categorization of these expressions by others. Both analyses established conclusively that these dynamic features are governed by specific and essential rules and cannot be eliminated from a serious discussion of ASL facial expressions.

Categorization studies of static facial expressions have found an accuracy advantage for deaf signers in categorizing sign language facial expressions (Campbell, Woll, Benson, & Wallace, 1999) and an increased ability of deaf signers to appreciate and recognize subtle changes in faces manipulated from the Benton Test of Face Recognition (McCullough & Emmorey, 1997). While this last study did not use ASL facial expressions, it is clearly relevant to the discussion of whether deaf signers can reliably discriminate between emotional and grammatical facial expressions that are often differentiated by only minor feature variations and whether they can do so better than hearing non-signers with less exposure to expressions taken from sign language.

Using naturally occurring ASL facial expressions, but also only in static representation, Corina (1989) found that deaf signers in a laterality paradigm showed visual field differences for grammatical versus emotional facial expressions, while hearing non-signers did not. These results are partially explained by his hypothesis that ASL grammatical faces (including wh-question face) are unfamiliar to hearing non-signers and therefore processed differently by these two subject groups. However, studies have shown that speakers of American English use a great variety of facial displays occurring both with and without spoken utterances (Chovil, 1991). The most common facial displays were found to be syntactic in nature and most often associated with specific grammatical constructs, such as questions. These results are supported by evidence from facial expression coding of English speakers' faces, indicating that questions are frequently marked by either eye brow raises or lowered brows (Ekman, 1979). Since ASL yes/no question faces involved raised brows, and wh-question faces contain lowered brows, it becomes reasonable to hypothesize that hearing non-signers, given the appropriate stimulus presentation, may well be able to access the meaning of ASL question faces and differentiate them from the more general conversational and emotional expression included in this study. Taking into account evidence that dynamic properties of facial expressions affect the way we recognize different expression types (Kamachi et al., 2001), and may even recruit a different set of neural substrates than static facial expressions (Adolphs, Damasio, Tranel, & Damasio, 1996; Kilts, Egan, Gideon, Ely, & Hoffman, 2003), it becomes crucial to revisit the recognition and categorization of facial expressions using dynamic stimuli, especially in light of recent findings that no individual snapshot out of a dynamic facial expression video contains the apex of more than one facial feature movement at a time (Grossman & Kegl, 2006). This lack of a single-image expression apex makes it clear that photographs of facial expressions portray, at best, a subjective ''maximum expression,'' but are not truly representative of the peak expression of all facial features involved. While it is certainly true that deaf signers can more readily discern small feature differences in facial expressions and are expected to be more familiar with the linguistic content of ASL grammatical facial expression, the question arises as to whether the established deaf advantage for recognizing and categorizing ASL facial expressions is maintained when we present the full range of features for these expressions, including their essential dynamic aspects.

The aim of this study is to determine whether hearing non-signers and deaf signers differ in their ability to distinguish between emotional and ASL-linguistic or conversational facial expressions based on video clip presentation as opposed to static images of several expressions in each category. We assess whether preserving the dynamic components of facial expression stimuli affects the participants' abilities to distinguish between superficially similar expressions.

Method

Stimuli

The stimuli used in this study consist of six types of facial expressions: *neutral*, *angry*, *surprise*, *quizzical*, *y/n question*, and *wh-question*. These stimulus types were chosen specifically because they contain feature similarities, but express very different content. Quizzical, angry and wh-question all exhibit furrowed brows and squinted eyes, while yes/no questions and surprise display widened eyes and raised brows.

Stimulus Generation

Stimuli were generated by a seventh generation deaf signer, JCG, who uses ASL as his primary means of communication in day-to-day activities at home, work, and school. JCG is a teacher of deaf children, an ASL instructor, and has considerable experience signing for a video camera. We used a single signer to create these stimuli for the sake of consistency of expressiveness across samples. A detailed analysis of JCG's facial expressions documented consistently more subtle expressions than those of a second signer (PLT) who was asked to produce the same type and quantity of stimuli. However, specific coding and analysis of all elicited facial expressions revealed a remarkable consistency between signers of the dynamic components that differentiate emotional and grammatical expressions (Grossman & Kegl, 2006). We therefore felt comfortable using JCG as the single expressor for the creation of these stimuli. The full list of stimulus sentences selected in collaboration with JCG is shown in English and ASL gloss in Table 1. The initial list of sentences was created using the following criteria: the stimuli had to be of approximately equal duration when signed, equally believable in all emotional and grammatical variations, and natural in everyday conversation.

In addition to those basic criteria and based on feedback we received from deaf participants of a pilot study, we eliminated all grammatical constructs that would require secondary facial expressions (e.g., negation, conditionals, or topic) besides those specifically targeted. This was done to ensure that the target facial expression was the only significant expression in the stimulus. All sentences were reviewed and approved by JCG in order to ensure that they were natural and comfortable for him to sign.

Table 1 List of stimulus sentences in ASL gloss and English translation

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

Wh-questions had a wh-word added to the end of the sentence

Quizzical sentences had "ME WONDER" added to the end of the sentence

Stimulus Recording

After the complete set of stimuli was transcribed on paper, we taped JCG's performance of the 6 versions (neutral, angry, surprise, quizzical, yes/no question and wh-question) for each of the 20 sentences. The stimuli for this study were recorded on a camera trained specifically on JCG's face, capturing his facial expressions exclusive of hand and arm movements. Two additional cameras captured JCG's torso exclusive of the face, and his entire body, respectively. The torso and whole-body videos were not used in the generation of these stimuli, but enabled us to corroborate the validity of each facial expression based on the corresponding manual component.

A crucial aspect of filming was to ensure that JCG's hands did not intrude into the 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 very difficult to establish a natural sample of ASL sentences excluding all signs produced at or near the face, the solution was to instruct JCG 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 both affective and grammatical facial expressions. The video clips were also monitored closely to ensure that there was no excessive mouthing of words that might have tempted participants to try and lip-read the meaning of the sentence during the experiment. We also recorded JCG signing the instructions to the task and a short concluding statement designed to thank the participants for their participation after completion of the task.

Video Editing, Stimulus Sequencing and Coding

After recording each of the 120 stimuli multiple times, the videotape was reviewed in order to select the best iteration of each stimulus. The 120 selected sentences were then coded in detail, using SignStream™, a multi-level database program developed in a collaboration among Boston University, Rutgers, the State University of New Jersey, and Dartmouth College (Neidle et al., 1997, 1998, 1999). In contrast to FACS, SignStream™ is designed to track and record the movement of features (e.g., hands, arms, eyebrows, or whole body movement), not muscle groups, in a frame-by-frame video analysis based on QuickTime files. Each individual user can adjust the types of features and coding definitions used to suit the needs of the videos being coded.[1] This allowed us to ensure that only stimuli that exhibited the facial feature movements expected for each expression type were included in this study (e.g., productions of wh-question faces that did not contain the expression-typical lowered brows and squinted eyes were eliminated from the stimulus pool).

The clips were then edited into 4 separately calculated counterbalanced sequences of the 120 expressions. Each sequence of the same 120 sentences was designed to maximize distribution of the different types of expressions and minimize

[1] For a detailed description of the SignStream™ analysis for these stimuli, see Grossman and Kegl (2006).

repetition of patterns. The 4 sequences were edited to contain all 120 facial expressions without repetition and were copied onto high-quality VHS tapes. Each presentation of an expression (one "clip") was preceded and followed by a blank screen with a fixation cross in the center. The tapes were edited to ensure an interval of exactly 14 s between the onsets of two subsequent clips. Participants in the pilot study deemed this interval sufficient to categorize each facial expression.

Participants and Procedure

The stimuli were shown to a group of 24 hearing participants and 18 deaf participants. Participants were recruited through e-mail notices or word of mouth/hand. All hearing participants were pre-screened to make sure they had no knowledge of ASL beyond basic fingerspelling. Only participants who were native English speakers and had grown up in the United States were included in the study. All deaf participants were pre-screened to ensure they were fluent signers who use ASL as their primary means of communication. Eight of those deaf participants were native signers, born into deaf families who use ASL to communicate.[2] The total number of participants included in the analyses for this study was 23 hearing and 15 deaf participants.

Both subject groups were given a detailed introduction to the task, written for the hearing participants, and in ASL on videotape for the deaf participants. Hearing participants were also shown parts of the ASL introduction to provide them with a facial expression baseline for JCG. Informed consent was obtained for all participants, in accordance with the study protocol approved by the Institutional Review Boards of Boston University and the University of Southern Maine, and all participants were reimbursed for their time. Participants were seated comfortably in front of a TV screen, with pre-printed response sheets on clipboards or tables in front of them. Participants either viewed the videotape alone or in small groups. The four different sequences of stimuli were used in an alternating pattern for each subject or subject group. Participants marked their category decisions on a response sheet containing all possible response categories for each stimulus presentation. The response categories were defined (e.g., *question* means a person is asking a question, *quizzical* means a person is saying a sentence that could start with "I wonder if...") and participants verified that they understood the differences among all stimulus categories prior to commencing the study.

The response sheets were identical for both subject groups, with one exception. The hearing participants were asked to determine simply if a stimulus expressed a question of any type. They were not asked to make a distinction between the two question types (wh-question and y/n question), since we did not want them to have to make syntactic distinctions in a language that was not familiar to them. In the pilot study, deaf participants received the same response sheet as the hearing participants, but quickly complained that they felt at a loss to classify the two different types of ASL question faces because they were given only one response category for both

[2] After completion of the task, one hearing subject's data were eliminated from the sample because she indicated belatedly that she was not a native English speaker. Two deaf participants were disqualified prior to completing the study. One was disqualified because she was unable to decide on how to categorize any of the stimuli, the other because his vision was so poor that he was not able to visually distinguish between any of the facial expressions presented. One additional deaf subject was excluded after completion of the study, because he consistently checked off more than one response category for each stimulus presentation, making it impossible to interpret his data.

question stimulus types. Since they could recognize that the stimuli contained two different types of questions, they tried to reflect that difference by checking off two different response categories for the ASL question stimuli. This resulted in participants over-selecting the *quizzical* or *other* response categories for ASL question stimuli because those two response categories were physically adjacent to the *question* response category on the check sheet. According to deaf participant feedback from the pilot study, dividing their responses in this way allowed them to feel that they were distinguishing between the two types of question face stimuli. Based on these comments, the response sheets for deaf participants were modified to reflect both types of ASL question faces and thus include seven response categories by splitting the *question* category in two, namely *wh-question* and *y/n question*. Both sets of response sheets included a 5-point confidence scale adjacent to the response line for each stimulus, allowing participants to mark both their category selection and their confidence about that selection on a single line. After completion of the study, participants were debriefed about their experience, thanked for their participation, and dismissed.

Results

Category Judgments

All participants' category decisions were collected and analyzed. The number of correct responses to each stimulus type (out of 20 stimuli within each type) was calculated for every subject. Analyses showed that neither subject group exhibited improvement due to learning over the course of the study. We also established that no individual stimulus caused significantly more error responses than others, indicating that the data were not skewed by a few inappropriate stimuli. The data for all participants were grouped into two separate averages for deaf and hearing with their respective standard deviations representing the variance between participants. To allow for group comparisons, the deaf participants' results were recalculated so that either question response category was counted as correct for either type of question stimulus. This process equalized the two subject groups by creating the same number of response categories (six) for both, therefore making it possible to compare them directly. Looking at the deaf participants' responses for ASL question stimuli before and after collapsing the response categories, it was clear that participants did not significantly choose the wrong question category for ASL question stimuli (i.e., selecting wh-question for the presentation of a y/n question and vice versa). It was therefore possible to collapse those two response categories into one.

Analysis of Variance

Overall accuracy means for both subject groups are listed in Table 2. The accuracy scores for deaf and hearing participants were analyzed in a 2×6 mixed two way analysis of variance (ANOVA) with stimulus type as the within-participants factor. The sphericity assumption was met. Results of that analysis indicate a main effect for subject group, $F(1, 36) = 8.44$, $p < .01$, with the accuracy mean for deaf participants ($M_D = 10.01$) being significantly *lower* than the mean for hearing participants ($M_H = 11.94$). There was also a main effect for stimulus type, $F(5, 180) = 30.16$,

Table 2 Means and standard deviations for emotion detection accuracy scores

Stimulus type	Mean accuracy (*SD*)	
	Deaf ($N = 15$)	Hearing ($N = 23$)
Neutral	12.2 (4.41)	15.87 (2.7)
Angry	15.4 (2.97)	17.13 (2.51)
Surprise	11.2 (3.53)	11.65 (3.1)
Quizzical	6.93 (5.18)	7.04 (4.06)
Yes/No	6.33 (6.55)	9.96 (5.12)
Wh	8.0 (5.21)	10.0 (3.83)

$p < .001$, with angry stimuli reaching the highest accuracy scores ($M = 16.27$), followed by neutral ($M = 14.04$), surprise ($M = 11.43$), wh-question ($M = 9.0$), y/n question ($M = 8.15$), and finally quizzical ($M = 6.99$). There was no significant subject group by stimulus type interaction, $F(5, 180) = 1.32$, $p > .2$. Post hoc comparisons between participants were made using Tukey's HSD test with p set at .05. The hearing participants were significantly more accurate than the deaf on neutral ($M_{\mathrm{H}} = 15.87$, $M_{\mathrm{D}} = 12.20$), angry ($M_{\mathrm{H}} = 17.13$, $M_{\mathrm{D}} = 15.40$), y/n question ($M_{\mathrm{H}} = 9.96$, $M_{\mathrm{D}} = 6.33$), and wh-question stimuli ($M_{\mathrm{H}} = 10.00$, $M_{\mathrm{D}} = 8.00$). All accuracy means and significances for both subject groups are listed in Table 2.

A chi-square analysis was conducted in order to establish whether the accuracy scores of both subject groups were above chance level. The hearing participants achieved accuracy levels significantly above chance for all stimulus categories, with all $\chi^2(1) \geq 9.78$, all $p < .01$. The deaf participants reached accuracy scores above chance for neutral, angry, surprise, and quizzical stimuli, with all $\chi^2(1) \geq 5.4$, all $p < .05$, but were *not* significantly different from chance for wh-question stimuli ($\chi^2(1) = 3.27$ $p > .07$) and y/n question ($\chi^2(1) = 0.67$, $p > .7$).

Error Analysis

In addition to the accuracy scores, we also established the number of hits—correct and incorrect—in all response categories for each stimulus type, thus creating an error analysis for each facial expression type. This latter analysis enables us to see whether the errors made by the deaf and hearing cohorts show specific patterns, such as systematically mistaking wh-questions for angry or quizzical facial expressions, the expressions that look most like each other (see Table 3). The error analysis shows that y/n questions were most often misidentified as surprise (the two expressions sharing raised brows and widened eyes), while wh-questions were most frequently mislabeled as quizzical, followed by angry (the three expressions sharing lowered brows and squinted eyes).

Confidence Ratings

The confidence scores for 14 deaf and 22 hearing participants[3] were analyzed in a 2×6 mixed two way ANOVA with stimulus type as the within-participants factor. The sphericity assumption was not met so the Greenhouse-Geisser correction was

[3] One each of the 15 deaf participants and 23 hearing participants included in the accuracy analyses did not mark any confidence rating.

Table 3 Confusion matrix for emotion detection accuracy and error responses

Response						
Stimulus	Neutral	Angry	Surprise	Quizzical	Question	Other
Neutral	*12.2***	0.6	0.9**	2.7*	2.8**	0.8
	*15.9***	1.2	0.1**	1.2*	0.3**	1.2
Angry	0.3	*15.4*	0.5	1.3	2.1	0.5
	0.3	*17.1*	0.3	0.7	1.2	1.2
Surprise	1	1.2	*11.1*	2.1	3.7	0**
	1.1	1.5	*11.7*	1.1	3.3	1.2**
Quizzical	4.8	0.8**	2.2**	*6.9*	4.1	1*
	5.3	2.3**	0.5**	*7*	1.9	2.9*
Yes/No	2.1	0.1	6.5	2.7	*7.9*	0.5
	2.2	0.2	5.8	1.7	*10*	0.1
Wh	0.4	3.2	0.9	4.8	*9.8*	0.8
	0.3	3.3	0.4	4.9	*10*	1

* $p < 0.05$

** $p < 0.01$

Significances refer to differences with other response categories for the same stimulus type

Top number represents deaf average accuracy (out of 20).

Bottom number represents hearing average accuracy (out of 20)

applied. Results of that analysis indicate a main effect for stimulus type, $F(3.9, 132.8) = 7.99$, $p < .001$ as well as a main effect for subject group $F(1, 34) = 10.24$, $p < .01$, and a subject by group interaction $F(3.9, 132.8) = 3.84$, $p < .01$. Overall confidence means and standard deviations for both subject groups are listed in Table 4. Post hoc comparisons between participants were made using Tukey's HSD test with p set at .05. The deaf participants were significantly more confident than the hearing group for angry ($M_D = 4.12$, $M_H = 3.66$), surprise ($M_D = 3.87$, $M_H = 3.40$), quizzical ($M_D = 3.79$, $M_H = 3.02$), y/n question ($M_D = 3.96$, $M_H = 3.35$), and wh-question stimuli ($M_D = 3.91$, $M_H = 3.22$), but not for neutral ($M_D = 3.74$, $M_H = 3.56$).

Correlation between Confidence and Accuracy

A series of two-tailed Pearson correlations were calculated in order to determine the association between accuracy scores and confidence ratings in each subject group. Results show no significant correlations between confidence and accuracy in either subject group for any stimulus type (all $r \leq -0.22$, all $p \geq .33$).

Table 4 Means and standard deviations for confidence scores

Stimulus type	Mean confidence (*SD*)	
	Deaf ($N = 14$)	Hearing ($N = 22$)
Neutral	3.74 (.66)	3.56 (.62)
Angry	4.11 (.61)	3.66 (.50)
Surprise	3.87 (.67)	3.40 (.39)
Quizzical	3.79 (.48)	3.02 (.66)
Y/N question	3.96 (.65)	3.35 (.49)
Wh question	3.91 (.63)	3.22 (.48)

Discussion

Accuracy Scores

The most striking accuracy result is that the deaf group exhibited lower accuracy scores than the hearing cohort. One possible explanation for this is connected to both the long exposure time to the stimulus and the deaf participants' increased knowledge of facial complexity. Both Wilson and Schooler (1991), and Edwards (1998) found that participants who were given too much time to consider their choices often scored lower on accuracy. In the context of the present study, both subject groups were given a relatively long time to consider their options (a 14-s-interval chosen as comfortable by participants in the pilot study), but the deaf group had more information on facial expressions at their disposal to influence their decision. It is possible that the deaf participants reduced their accuracy scores even further than the hearing participants by ''over-thinking'' what should have been an intuitive choice.

The deaf group's lower accuracy results are the most puzzling in the two ASL questions, for which deaf participants were not only less accurate than the hearing group, but not even significantly different from chance, despite their supposed greater familiarity with these language-based facial expressions. This result stands in contrast to the deaf signers' advantage for ASL facial expressions found in studies using static images.

A possible explanation for this counter-intuitive finding is that the deaf signers were faced with a more difficult task by having to categorize the two ASL question faces separately, rather than in one collective ''question'' response category. We gave the deaf signers two separate responses categories (wh-question and y/n question) because deaf pilot participants requested the opportunity to reflect the difference between the two stimulus types on their response sheet. It was obvious to the pilot participants that the stimuli included two different types of ASL question faces (y/n and wh-) and they were not satisfied with having only a single response category (*question*) to reflect these two stimulus types. However, it is possible that this increased specificity in categorization of ASL question expressions required of the deaf signers ultimately served to decrease their accuracy for these stimulus types.

JCG's facial expressions were intentionally kept at a low level of intensity in order to avoid the unnatural ''overdrawn'' presentation of facial expressions participants criticized in a pilot study using different stimuli. Palermo and Coltheart (2004) indicated a significant correlation between recognition accuracy and facial expression intensity ratings and it is possible that the laid-back nature of JCG's facial expressions contributed to reducing the accuracy levels for all participants and stimulus categories. However, the great variability in the accuracy scores of individual participants, ranging from 0% to 100% correct in the deaf, and 5 to 100% correct in the hearing, argues against the existence of a low ceiling effect in these data. Furthermore, data not included in the group averages show that JCG's deaf wife scored an average of 91% correct for all stimulus types, ranging from 80% for wh-questions, to 100% for neutral expressions. These results speak against the assumption that the stimuli were too subtle to be interpreted correctly and that ceiling for accuracy was below 80% correct. While JCG's stimuli were clearly conveying the target expression to his wife, they proved somewhat more challenging for the participants who were less familiar with him. The data also reinforce findings by Noller and Ruzzene (1991) and Sabatelli, Buck, and Dreyer (1982) that romantic partners are better at reading each others' facial cues than strangers.

Both subject groups achieved the highest accuracy scores for angry expressions, followed by neutral and then surprise. This pattern agrees with established data that indicates different levels of accuracy and response times for various facial expression types, with anger being the most easily identified and surprise being the least easily identified among expressions of basic emotion (Ekman, 1982; Wallbott & Sherer, 1986).

Both groups also categorized the majority of quizzical expressions correctly, with their most common error being to categorize them as neutral. A detailed facial-coding analysis (Grossman, 2001; Grossman & Kegl, 2006), using a novel modification on the SignStream™ software package, demonstrates that quizzical expressions are defined by distinct dynamic facial feature events that do not exist in neutral expressions, indicating that this pattern of confusion is not based on superficial similarities between the two expressions. The expressions that do share feature similarities with quizzical faces are angry and wh-question, but although wh-question faces were frequently misinterpreted as angry or quizzical by both subject groups, the reverse was not true. The error analysis for ASL question faces shows that y/n questions were most often misidentified as surprise (the two expressions sharing raised brows and widened eyes), while wh-questions were most frequently mislabeled as quizzical, followed by angry (the three expressions sharing lowered brows and squinted eyes). In general, we can ascertain that all participants showed a secondary tendency to misinterpret facial expressions according to superficial feature similarities, such as eye and brow involvement.

Overall, there appears to be no advantage for deaf signers during this facial expression categorization task, despite their greater experience with ASL question faces, which brings up the question of why the hearing participants were so easily able to identify the two ASL expressions as question faces. After completion of the task, all participants were asked to report on their perception of the task. Several hearing participants reported having seen facial expressions that looked like the signer was asking a stereotypic "whaaat?" question. The expression the participants modeled when saying this, was one of general incredulity, so ubiquitous in colloquial American English that it can be understood even without verbalization. Imagine the face of a teenager while stating: "He's telling me this whole story and I'm looking at him, like ["whaaat"-face]." That facial expression is created by furrowing the brows, squinting the eyes, raising the upper lip and often accompanied by a head shake, showing at least superficial similarity to the grammatical wh-question face in ASL in both dynamic and facial feature involvement (Grossman, 2001; Grossman & Kegl, 2006). The meaning of these two expressions cannot be exactly equated. The ASL wh-question face is a required component of the language in any context, while the "whaaat"-face is highly colloquial in English and does not occur with any regularity during simple wh-questions. It seems to express a more general state of incredulity or doubt than an actual wh-question. Nevertheless, if hearing non-signers had to put a word to that expression, it would be "what?" Based on this subject feedback and the evidence in the data, we hypothesize that the dynamic and facial feature components of this familiar colloquial template allow hearing non-signers to recognize ASL wh-question faces as expressing a question.

Hearing participants did not specifically report mapping ASL y/n question faces to an English template, but the standard English question face (raised eyebrows, widened eyes) does fill that role. As with the ‘‘whaaat’’-face, the general question face is not a required element of a grammatical English y/n question, but it occurs frequently during natural conversation and can also be used without any verbal accompaniment (Chovil, 1991). If one person is telling a story and pauses before the punch-line, the conversation partner may well look at the speaker with raised eyebrows, widened eyes, and frequently a lowering of the chin, in a facial expression that strongly resembles the ASL y/n question face. That expression in this context is akin to asking, ‘‘So what happened?’’ Although we do not yet know the specific dynamic properties of this English question face, we do know that it superficially resembles the ASL y/n question face and therefore may allow non-signers to recognize the question content of this stimulus type. The accuracy levels achieved by the hearing non-signers for categorization of ASL y/n question faces certainly indicate that hearing participants were able to recognize the meaning of the ASL question expressions.

None of these data suggest that ASL question faces are not an integral and standardized part of the language, but rather indicate that stereotypical versions of these expressions also exist in the North American spoken language culture and can therefore be recognized by hearing non-signers. This similarity is reasonable considering that both languages evolved in the same social and cultural environment. The existence of non-random head movements in hearing Americans as a potential source for more standardized head movements in ASL has been suggested before (McClave, 2000, 2001). The fact that the hearing participants in this study performed so well in the task of categorizing ASL question faces seems to support that hypothesis. However, it is important to emphasize that facial expressions used in spoken English are neither required, nor necessarily standardized, while ASL question facial expressions are a required, standardized component of a complete, grammatical ASL phrase to which specific rules of onset and offset apply (Grossman, 2001).

The findings presented in this study, which contradict prior findings of a deaf accuracy advantage for sign language facial expressions, depend critically upon the use of a medium (video) that preserves the dynamic changes occurring during the development of a facial expression, as opposed to the more commonly used still photographs. Grossman (2001), and Grossman and Kegl (2006) emphasize the importance of preserving the dynamic elements of facial expressions. Different types of facial expressions that share superficial features, such as wh-question and quizzical, or y/n question and surprise, were examined in detail for dynamic facial features. Even though two different expressions may share a lowering of the eyebrows or narrowing of the eyes, the dynamic development of these movements, i.e., the speed and modulation of the movement over time, as well as unique dynamic identifiers found in only one of the expression types, can be used to distinguish between them. Importantly, the analysis revealed that there was no specific point in time or individual video frame for any expression type during which all features were at their maximum expression. This clearly shows that still photographs of a ‘‘peak expression’’ do not, in fact, represent the peak of the overall facial expression, but rather the apex of only one of the several simultaneously evolving facial features at a time. Using still photographs therefore does not allow participants to tap into all the information necessary to decode facial expressions. Only the presentation of video clips preserves all unique identifiers of individual expression types, which may significantly affect the ability of participants to categorize emotional and grammatical facial expressions. Our study found reduced accuracy for deaf participants' categorizations of dynamic ASL question faces compared to hearing non-signers, as opposed to a deaf signer advantage in categorizations of static ASL faces documented in the literature. This discrepancy should be investigated further, preferably by using both static and dynamic stimuli in the same experimental design.

Confidence Rating

Even though the hearing participants were significantly more accurate for all expression types, the deaf cohort was consistently more confident of their selections than their hearing counterparts. This discrepancy fits into existing research showing a merely tenuous connection between accuracy and confidence (Ames & Kammrath, 2004). However, there is also evidence to suggest that increased exposure to a stimulus results in increased confidence (Robert, Deffenbacher, & Brigham, 1987). It is possible that the deaf signers' long experience with these types of stimuli feeds into the same mechanism and enables them to feel more confidence about their selections than the hearing participants, despite their poor accuracy performance.

Deaf signers maintain more and longer eye contact during conversation than hearing speakers of English (Kegl, Cohen, & Poizner, 1999; Swisher, 1991), indicating greater focus on their conversation partner's face than hearing non-signers usually show. Deaf signers must maintain eye contact with each other during conversation to extract all linguistic information, manual and facial. Even though ASL is a visuo-spatial language that conveys most of its semantic content through the hands and arms, signers focus on each other's faces during conversation and do not follow their conversational partner's hand movements with their eyes (Siple, 1978). Hearing non-signers, on the other hand, can follow spoken conversations without focusing on the speaker's face, and look to the face primarily for supplemental emotional and discourse information during a language interchange. This experience of paying attention to the face during conversation is what deaf signers can rely on during the performance of this task. It may very well be that this is what drives their increased confidence during category selections, especially considering that their confidence levels are higher for all stimulus categories, not just the ASL question faces.

Looking at confidence scores across stimuli, we find that the deaf participants were the least sure of neutral stimuli, while the hearing cohort scored their second highest confidence ratings for those stimuli. This result resonates with the complexities of facial expressions that deaf signers deal with on a daily basis. Deaf signers of ASL look to the face for a multitude of syntactic, semantic, and prosodic information. In order to clearly delineate the scope of this study, we eliminated all extraneous facial expressions from JCG's productions by ensuring that the underlying sentences did not require any facial expression other than the one we targeted. In addition, our specific instructions to the participants were not to look at neutral as "expressionless," but merely as not displaying any of the other target expressions. However, taking into account the deaf signers' vast experience with the multitude of facial expressions commonly exhibited during normal conversation, it is possible that, despite our instructions, they were simply less comfortable than the hearing to call any facial expressions neutral.

It is interesting to note that the deaf participants reached their second and third highest confidence levels for y/n questions and wh-questions, despite not performing above chance for accuracy on these expressions. Conversely, the hearing subject group scored significantly above chance for accuracy in categorizing the two ASL question faces, but their confidence levels for them received the two lowest scores. It is possible that familiarity with these stimuli was reflected only in the confidence levels, but not in the accuracy scores, a result that should be investigated further.

Conclusion

The hearing group showed a significantly higher level of accuracy than the deaf cohort for all stimulus types. The lack of a deaf accuracy advantage for ASL question faces shown in previous studies using static images leads us to believe that the preserved dynamic components of the stimuli in this study have affected the ability of hearing non-signers to access the meaning of these facial expressions. Facial displays used in American English dialogue may serve as templates to assist hearing non-signers in understanding the ASL question faces presented in this study.

Confidence measures made at the same time as the categorizations revealed a reverse pattern, with the deaf participants expressing significantly more confidence about their choices than the hearing. This increased confidence of the deaf participants despite lower accuracy scores may be a result of their greater experience with the stimuli. Further studies of dynamically presented facial expressions should be conducted, especially in conjunction with presentations of static facial expression stimuli, in order to establish the specific way in which the dynamic properties of English and ASL facial expressions inform recognition and categorization.

Acknowledgments The data reported here were collected in part under NSF Grant SBR-929065. The authors would like to thank Anthony Barrand, Rhyannon Bemis, David Black, Christine Connolly, Wayne Cowart, Rebecca Damon, Nat Ives, Bob Joseph, Leila Monaghan, MaryAnn Nilsson, Robert Savoy, Romy Spitz, and Helen Tager-Flusberg for their assistance in this study and the preparation of this manuscript. We also thank the Deaf signer (JCG) who produced the stimuli and the USM Media Studies Program for use of their equipment.

References

Aarons, D. (1996). Topics and topicalization in American Sign Language. *Stellenbosch Papers in Linguistics, 30*, 65–106.

Adolphs, R., Damasio, H., Tranel, D., & Damasio, A. R. (1996). Cortical systems for the recognition of emotion in facial expressions. *The Journal of Neuroscience, 16*(23), 7678–7687.

Ames, D. R., & Kammrath, L. K. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior, 28*(3), 187–209.

Bahan, B. (1996). *Non-manual realization of agreement in American Sign Language*. Boston: Unpublished doctoral dissertation, Boston University.

Baker-Shenk, C. (1983). *A micro-analysis of the nonmanual component of questions in American Sign Language*. Berkeley: Unpublished doctoral dissertation, University of California.

Baker-Shenk, C. (1986). Factors affecting the form of question signals in American Sign Language. In: D. Sankoff (Ed.), *Diversity and diachrony* (pp. 407–414). Amsterdam, Philadelphia: John Benjamins.

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 373–379.

Bassili, J. N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology, 37*, 2040–2058.

Campbell, R., Woll, B., Benson, R. J., & Wallace, S. B. (1999). Categorical perception of face actions: Their role in sign language and in communicative facial displays. *The Quarterly Journal of Experimental Psychology, Part A, 52*(1), 67–95.

Chovil, N. (1991). Discourse-oriented facial displays in conversation. *Research on Language and Social Interaction, 25*, 163–194.

Corina, D. P. (1989). Recognition of affective and noncanonical linguistic facial expressions in hearing and deaf participants. *Brain and Cognition, 9*(2), 227–237.

Edwards, K. (1998). The face of time: Temporal cues in expressions of emotion. *Psychological Science, 9*(4), 270–276.

Ekman, P. (1979). About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology* (pp. 169–202). Cambridge: Cambridge University Press.

Ekman, P. (Ed.). (1982). *Emotion in the human face*. Cambridge, England: Cambridge University Press.

Ekman, P. (1984). Expression and the nature of emotion. In: K. R. Sherer, P. Ekman (Eds.), *Approaches to emotion* (pp. 319–343). Erlbaum: Hillsdale, NJ.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face. A guide to recognizing emotions from clues*. Englewood Cliffs, NJ: Prentice-Hall.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.

Emmorey K. (1995). Processing the dynamic visual–spatial morphology of signed languages. In: L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 29–54). Erlbaum: Hillsdale, NJ.

Grossman, R., & Kegl, J. (2006). To capture a face: A novel technique for the analysis and quantification of facial expressions in American Sign Language. *Sign Language Studies. 6*(3), 273–305.

Grossman, R. (2001). *Dynamic facial expressions in American Sign Language: Behavioral, neuroimaging, and facial coding analyses for deaf and hearing participants*. Boston, MA: Unpublished doctoral dissertation, Boston University.

Hoza, J., Neidle, C., MacLaughlin, D., Kegl, J.A., & Bahan, B. (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 No. 4). Boston, MA: Boston University.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception, 30*, 875–887.

Kegl, J., Cohen, H., & Poizner, H. (1999). Articulatory consequences of Parkinson's Disease: Perspectives from two modalities. *Brain and cognition, 40*(2), 355–386.

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage, 18*, 156–168.

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics, 7*(32), 855–878.

McClave, E. (2001). The relationship between spontaneous gestures of the hearing and American Sign Language. *Gesture, 1*(1), 51–72.

McCullough, S., & Emmorey, K. (1997). Face processing by deaf ASL signers: Evidence for expertise in distinguished local features. *Journal of Deaf Studies and Deaf Education, 2*(4), 212–222.

MacLaughlin, D. (1997). *The structure of determiner phrases: Evidence from American Sign Language*. Boston, MA: Unpublished doctoral dissertation, Boston University.

Neidle, C., MacLaughlin, D., Bahan, B., & Kegl, J. (1996). Non-manual correlates of syntactic agreement in American Sign Language (American Sign Language Linguistic Research Project Report No. 2). Boston, MA: Boston University.

Neidle, C., MacLaughlin, D., Bahan, B., Lee, R., & Kegl, J. (1997). The SignStream^{TM} Project (American Sign Language Linguistic Research Project Report No. 5). Boston, MA: Boston University.

Neidle, C., Foelsche, O., Kegl, J., MacLaughlin, D., Bahan, B., & Greenfield, D. (1997, 1998, 1999). SignStream^{TM} Software, Version 1.5 [Computer software]. Trustees of Dartmouth College, Trustees of Boston University and Rutgers, the State University of New Jersey.

Noller, P., & Ruzzene, M. (1991). Communication in marriage: The influence of affect, cognition. In G. J. O. Fletcher & F. D. Fincham (Eds.), *Cognition in close relationships* (pp. 203–233). Erlbaum: Hillsdale, NJ.

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavioral Research Methods, Instruments, and Computers, 36*(4), 634–638.

Petronio, K., & Lillo-Martin, D. (1997). WH-movement and the position of spec-CP: Evidence from American Sign Language. *Language, 71*(1), 13–29.

Robert, K. B., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72*(4), 691–695.

Sabatelli, R. M., Buck, R., & Dreyer, A. (1982). Nonverbal communication accuracy in married couples: Relationship with marital complaints. *Journal of Personality and Social Psychology, 43*, 1088–1097.

Siple, P. (1978). Visual constraints for sign language communication. *Sign Language Studies, 19*, 95–110.

Swisher, M. V. (1991). Conversational interaction between deaf children and their hearing mothers: The role of visual attention. In P. F. Siple, & D. Susan (Eds.), *Theoretical issues in sign language research: Vol. 2. Psychology*. (pp. 111–134). Chicago/London: University of Chicago Press.

Wallbott, H. G., & Sherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51*(4) 690–699.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60*, 181–192.