Behavior Research Methods, Instruments, & Computers
2003, 35 (4), 590-598

Comparison of video- and EMG-based evaluations of the magnitude of children's emotion-modulated startle response

MARILYN J. ESSEX, H. HILL GOLDSMITH, NANCY A. SMIDER, and ISA DOLSKI
University of Wisconsin, Madison, Wisconsin

STEVEN K. SUTTON
University of Miami, Coral Gables, Florida

and

RICHARD J. DAVIDSON
University of Wisconsin, Madison, Wisconsin

We investigated the reliability and validity of a video-based method of measuring the magnitude of children's emotion-modulated startle response when electromyographic (EMG) measurement is not feasible. Thirty-one children between the ages of 4 and 7 years were videotaped while watching short video clips designed to elicit happiness or fear. Embedded in the audio track of the video clips were acoustic startle probes. A coding system was developed to quantify from the video record the strength of the eye-blink startle response to the probes. EMG measurement of the eye blink was obtained simultaneously. Intercoder reliability for the video coding was high (Cohen's $\kappa = .90$). The average within-subjects probe-by-probe correlation between the EMG- and video-based methods was .84. Group-level correlations between the methods were also strong, and there was some evidence of emotion modulation of the startle response with both the EMG- and the video-derived data. Although the video method cannot be used to assess the latency, probability, or duration of startle blinks, the findings indicate that it can serve as a valid proxy of EMG in the assessment of the magnitude of emotion-modulated startle in studies of children conducted outside of a laboratory setting, where traditional psychophysiological methods are not feasible.

The quest for objective measures of individual differences in emotional reactivity has been a major theme in both the developmental (e.g., Goldsmith, 1993) and the adult (e.g., Davidson, 1992) literature (see Davidson & Irwin, 1999, and Davidson & Sutton, 1995, for reviews). Particularly critical is the development of methods that are based on well understood brain mechanisms so that relations between behavioral and biological measures might be more fruitfully investigated. Of specific interest for developmental researchers is the establishment of objective measures of emotional reactivity that can extend the collection of data related to these principles of brain function to nonlaboratory studies as well as to other situations in which traditional physiological collection methodologies are not feasible or even possible.

In this report, we present empirical support for a method of video-scoring the startle response elicited by brief acoustic probes while children watch video clips designed to elicit positive (happy) and negative (fearful) emotion. The children were tested in a laboratory while simultaneous EMG and video measures of eye-blink magnitude in response to an acoustic startle probe were obtained.

Emotion-modulated startle is a very promising measure of individual differences in emotional reactivity, for several reasons. Its use in humans is based on an extensive corpus of data in animals, in which the detailed circuitry underlying the potentiation of the startle reflex during fear has been characterized (Davis, 1992). It is now known that forebrain structures modulate the brainstem startle circuit, which in turn provides the proximal control of the startle reflex. In humans, Lang and his colleagues have demonstrated that EMG measures of eye-blink magnitude from the orbicularis oculi region in response to a brief acoustic startle probe are larger when a subject is viewing negative pictures in comparison with

This research was supported by the Wisconsin Center for Affective Science (NIMH Grant P50-MH52354, Richard J. Davidson, Director), NIMH Grant MH43454 to R.J.D., NIMH Research Scientist Award K05-MH00875 to R.J.D., and by a grant from the John D. and Catherine T. MacArthur Foundation Research Network on Psychopathology and Development to M.J.E. The authors thank Darren Dottl for assistance with hardware and software; Rody Larson, Tracy Buechner, Stephanie McNeill, Ben Marion, and Jeffrey Armstrong for assistance in recruitment, testing, and behavioral coding; and Terry Kinney and Adrian Pederson for assistance with the stimulus tapes. Correspondence concerning this article should be addressed to M. J. Essex, Department of Psychiatry, University of Wisconsin–Madison, 6001 Research Park Blvd., Madison, WI 53719 (e-mail: mjessex@wisc.edu).

Copyright 2003 Psychonomic Society, Inc.

neutral and positive pictures. Eye-blink magnitude during viewing of positive pictures is typically smaller than that during viewing of neutral pictures (Lang, 1995; Lang, Bradley, & Cuthbert, 1990). It is postulated that these systematic valence-modulated effects on startle represent the influence of amygdala output on the brainstem startle circuitry. Using EMG measures of eye-blink magnitude, Cook, Hawk, Davis, and Stevenson (1991) demonstrated that subjects who rated higher in fear showed greater accentuation of startle response during a negative foreground stimulus than did low-fear subjects. Other investigators have observed systematic differences in emotion-modulated startle between other clinical groups hypothesized to have emotional deficits and normal controls (Patrick, Bradley, & Lang, 1993).

Although it has received only scant attention in the research literature, emotion-modulated startle can be used effectively in infants and children to provide an objective measure of emotional reactivity. For example, Balaban (1995) examined eye-blink magnitude using EMG measures in 5-month-old infants during the presentation of pictures of angry, happy, and neutral faces. She found the typical valence effect that has been found repeatedly in the adult literature. Blink magnitude during the presentation of the angry faces was significantly larger than during the presentation of the happy faces. And, in a sample of subjects aged 10 to 17 years, children of parents with a history of anxiety disorder had accentuated startle response magnitudes (Grillon, Dierker, & Merikangas, 1997). On the basis of these results, the startle reflex was advanced as a possible marker of vulnerability for the development of anxiety disorders later in life.

As these studies demonstrate, the EMG measure of blink magnitude provides a robust, easily quantified index of the strength of the eye-blink component of the startle reflex. This standard method, as well as others, such as the photocell technique, in which the closing eyelid interrupts a light beam (Flaten, 1993), is well suited for studies that involve laboratory visits. However, many large-scale developmental studies consist of home-based assessments in which physiological methods cannot be used. Thus, the present study was designed to develop a method of collecting information on the magnitude of children's emotion-modulated startle response that is feasible in studies outside of the traditional laboratory setting. We wished to develop a method of assessing emotion-modulated startle that could be derived from the video scoring of a child's eyes in response to acoustic startle probes presented during emotion-eliciting video clips. This approach is highly portable and easily administered in field settings, does not require specialized training for administration, and uses only ordinary video equipment for data collection. In addition, the startle probes are inserted on the audio track of the videotape so that no special apparatus is required for probe presentation. Because startle reactivity is captured on high-grade videotape, there is very good postprobe resolution of the magnitude of the eye-blink reaction. We are not proposing, however, that this method replace the more temporally and physiologically sensitive EMG measures that are feasible in laboratory settings.

We compared directly, in the same participants, EMG- and video-derived measures of reflex magnitude recorded simultaneously in response to brief acoustic probes. The probes were inserted on the audio track of video clips depicting positive (happiness-inducing) and negative (fear-inducing) animated scenes selected for presentation to young children between the ages of 4 and 7 years. Children were brought to the psychophysiology laboratory, where both EMG measures and videotaping of the face could be performed simultaneously. The utility of the video-based measure of reflex magnitude would be established by demonstrating that it can be coded from videotape with high intercoder reliability and that EMG-derived and video-based measures of blink magnitude in response to the identical startle probes are correlated, thus establishing the validity of the video coding. The utility of the specific administration protocol (e.g., the stimulus tape) would be supported by evidence of valence-specific modulation effects on the startle response.

METHOD

Participants

The participants were 31 children (18 boys, 13 girls) between the ages of 3.7 and 7.3 years (M = 5.7 years). The children's parents were paid $20 for their participation in a 1-h session. At the conclusion of the session, each child was allowed to choose a token reward for participation. All the children were Caucasian (n = 29) or African-American (n = 2), lived in the Madison, Wisconsin area, and had no known history of neurological disorder. All the parents provided written informed consent prior to the experiment.

Procedure

Each parent and child came to a waiting room in the psychophysiology laboratory. After the experimenter explained the session, the experimenter, child, and parent went into the testing room. Electrodes were attached to the child and impedances were checked (see EMG data acquisition below). Following electrode placement, the parent left the testing room and returned to the waiting room.

Video data acquisition and scoring. The video-based protocol was one that could be administered in a nonlaboratory field study. The child was seated on a 14-in. aluminum camping stool that was placed 36 in. from a 13-in. color video monitor (Sony TV/VCR Model KV13VM20). During the procedure, the child's face was videotaped using a Sony TR-83 8-mm analog camcorder, which captured 30 frames/sec (33.3 msec/frame) and recorded by progressive (noninterlaced) scan.[1] The camera was placed on a tripod as close as possible against the side of the TV monitor. The height of the tripod and the angle of the camera were adjusted to obtain a level, straight-on view of the child's face and eyes. The child's face and uppermost part of the body (from the top of the forehead to the very upper shoulders) were framed using the camera zoom. The time function on the camera was activated so that the videotape of each participant had a running time stamp. A camera operator sat silently behind the camera to ensure that the child's face would be captured even if the child shifted or moved. The camera operator was instructed never to look up from the camera or make any eye contact or conversation with the child.

The stimulus tape consisted of two positive (happiness-inducing) and two negative (fear-inducing) animated video clips. The clips

were extracted from four films that were selected for their emotion-eliciting potential and their appropriateness for presentation to young children. The clips were presented to the children in this order: *101 Dalmatians* (positive), *Lady and the Tramp* (positive), *The Secret of NIMH* (negative), and *The Land Before Time* (negative). The specific clip in each film was selected to maximize the likelihood of eliciting the desired emotion (e.g., the fear-inducing chase scene in *The Land Before Time*); the duration of the clips ranged from 50 to 124 sec. The order of the clips was invariant. In large field-based studies, the collection of startle data is likely to be only one of a number of activities and tasks designed to examine individual differences in socioemotional development. In such studies, the decision is often made to keep the order of tasks invariant across participants. Because the ultimate goal of this study was to establish a method that could be exported to large field-based studies, we elected not to counterbalance the order of presentation of the video clips. Moreover, the decision to present the positive clips first was made on the basis of two considerations. First, there is evidence of greater carryover effects of negative stimuli in comparison with positive stimuli (Davidson, Ekman, Saron, Senulis, & Friesen, 1990), and presenting the negative stimuli last worked against our hypothesis of greater startle magnitude in response to negative in comparison with positive stimuli because of the known habituation of the startle reflex over the course of an experimental session (see, e.g., Sutton, Davidson, Donzella, Irwin, & Dottl, 1997). Thus, any finding of greater startle magnitude during presentation of the negative stimuli than during that of the positive stimuli is likely to be an underestimation of the true effect size for this condition difference.

The audio track was presented binaurally through headphones (Sony Model MDRV100). Acoustic startle probes were inserted on the audio track of the videotapes. Startle probes (50 msec, 95-dB-SPL white noise burst with a nearly instantaneous rise time) were generated with a Coulbourn S81-02 noise generator and a Coulbourn S82-24 audio-mixer power amplifier. There were two probes per video clip. The intervals between the two probes varied between 26 and 47 sec. Probes were placed strategically within the clips in an attempt to maximize the effect of the specific emotion being elicited (e.g., immediately following a large and contagious grin by a Dalmatian). In all, a total of eight probes were presented during the four emotional video clips.

The acoustic characteristics of the postproduction stimulus tape were examined to determine the specific parameters of the stimulus tape as it was presented to the children via videotape. To do this, the audio output of the videotape was measured in two ways. First, a sound meter was used to determine the peak magnitude of the startle probes at the volume level at which they were presented to the children. The meter readings showed that the eight probes were virtually identical in absolute peak volume (range = 95.5–97 dB-SPL, M = 96.1 dB-SPL). Second, the entire audio track of the tape was saved as a .wav file using the Cool Edit Pro program (Syntrillium Software Corporation, Phoenix, AZ; www.syntrillium.com). The .wav file (in which magnitude is measured in dB-volume units or VU) allowed us to examine the average magnitude of the video soundtrack—that is, the background audio—in the 0.5 sec prior to the probes, the average difference between the background audio and the probe peak magnitude, the rise time of the probes, and other parameters of the videotape audio track. A review of the .wav file showed that in the 0.5 sec prior to the probes, the background audio was slightly louder for the negative clips (−4.1 dB-VU) than for the positive clips (−6.2 dB-VU).The average difference between the background audio and the probe peak magnitude was smaller for the negative clips (2.7 dB-VU) than for the positive clips (4.8 dB-VU). Visual inspection of the startle probes using both time and frequency domain representations of the .wav file indicated an instantaneous rise time for each probe. The duration of each of the eight probes was 46–47 msec. There was no evidence for degradation or distortion of the acoustic probes due to recording onto or playback from the audio track of the videotape.

For presentation of the stimulus tape during data collection, the audio track of the tape was run through a splitter with sound going simultaneously directly to the child via the headphones and to the microphone input of the video camera. This was necessary so that the videotape data would contain the exact occurrence of the probes. Prior to the presentation of the four video clips to the child, the experimenter listened and watched a short (30-sec) test clip to ensure the correct volume of the audio track and functioning of the video equipment.

Prior to the start of each of the four video clips, the experimenter gave the child a two-sentence synopsis of the plot leading up to the particular clip and reminded the child to watch carefully and to keep his or her hands in his or her lap. The experimenter was trained to use a tone of voice to help elicit the emotion of the video clip. For example, for the clip extracted from *101 Dalmatians*, the synopsis read to the child was the following: "The first cartoon is a happy one. The daddy dog is *so* happy because the mommy dog is having lots and lots of puppies." The headphones were then placed on the child's head, centered directly over the ears. Correct placement of the headphones over the ears was checked prior to each video clip. Because the child did not see the endings of the video clips, the experimenter assured the child at the end of the session that the negative video clips had happy endings and gave a brief synopsis of the resolution.

The magnitude of the eye blink was visually scored from videotape on a seven-point scale (from low to high). Viewing was done at full speed and then at slower speeds and frame by frame, until the rater was confident in her assessment of the magnitude of the eye blink. For each child, the rater noted the amount of eye showing when the child's eyes were at their normal level of openness while the child was watching the video clip just prior to presentation of the acoustic probe. The magnitude of the eye blink was then assessed in comparison with the baseline level of eye openness.

A score of 1 was assigned when the rater could see no muscle movement in response to the probe or when a trained rater had trouble *consistently* seeing an exceedingly small response. A score of 2 was assigned when the rater could consistently see a small eyelid flutter but the magnitude was not great enough to be considered a quarter blink. A score of 3 was assigned when the eyelid lowered approximately a quarter of the range for that child. A score of 4 was assigned for a half blink, a score of 5 was assigned for a 3/4 blink, and a score of 6 was assigned for a fully closed, soft eye blink (i.e., the lid completely closed, but there was no noticeable additional muscle tightening or "scrunch" around the eye). A score of 7 (full, hard blink) was assigned when there was visible muscle "scrunching" around the eye that accompanied the full blink. A blink was considered invalid if the probe occurred in the midst of a naturally occurring blink (in which the onset of the blink preceded the onset of the probe).

Training for video coding consisted of two steps. First, as an overview of the coding scheme and demonstration of what one is looking for, a videotape was created to illustrate a prototypic blink for each of the seven categories, and a manual containing descriptive explanations of the categories and the coding process was created. Second, an expert coder trained a second coder using practice tapes that the expert had previously coded. Once the second coder felt comfortable with the coding system and displayed good correspondence with the expert coder on the practice tapes (> 80% agreement), coding of the data commenced.

Determination of intercoder reliability was based on 30% of the cases. A stringent randomization and blinding process was used for these cases. Specifically, to establish intercoder reliability, a rating tape was created by randomizing the eight probes per participant, resulting in 88 individual coding segments. Each of the randomized

segments consisted of approximately 10 sec of video (8–10 sec prior to the probe and 2–3 sec after the probe). Two coders scored the tape independently. The first coder scored the tapes with sound and recorded the exact second at which the probe was presented. A second copy of this rating tape was made without a sound track. The second coder rated the tape without sound (i.e., blind to the positive/negative valence of the stimulus) using only the time mark provided by the first rater to indicate probe placement. After coding without sound, the rater reviewed the tape with sound to verify that, in fact, only probe-related blinks had been coded. Intercoder reliability (Cohen's κ) based on these 88 probes was .90. In no instance did the two coders differ by more than one scale point in their ratings. Discrepancies were reviewed and assigned scores by consensus. The remaining cases were scored, with sound, by consensus.

EMG data acquisition and reduction. The magnitude of the eye-blink response was recorded electromyographically via two 3-mm Ag–AgCl electrodes placed directly below the left eye (as per Vrana, Spence, & Lang, 1988). A third electrode was placed at the center of the forehead and was used as a ground.

EMG signals were amplified 30,000 times using Grass Model 12 amplifiers with bandpass filters set at 1 and 300 Hz and with the 60-Hz notch filter in. EMG signals were sent through a Rockland highpass filter set at 30 Hz. The signals were then integrated and rectified using a Coulbourn S76-01 contour-following integrator with a time constant set below 20 msec. All the signals were digitized at 1000 Hz on a 486 PC using SnapStream software (HEM Data Corp., Springfield, MI) and a 12-bit A/D board (Analogic Corp., Wakefield, MA). To calibrate the signals, signals of known frequency and magnitude were recorded prior to and following each session.

Eye-blink magnitude was quantified from the EMG signals using the following procedures. First, automatic peak and blink onset detection was performed on the integrated EMG response to each probe using an in-house software package. Each response was then reviewed by laboratory personnel. Any eye blinks were removed if they were associated with excessive noise during a 50-msec prestartle baseline period (e.g., blinks, unusually high amounts of integrated EMG during baseline) or because the onset of the integrated EMG eye-blink reflex began less than 20 msec following the acoustic probe onset. Eye-blink reflex magnitudes (in μV) were calculated by subtracting the amount of integrated EMG at reflex onset from the peak magnitude (maximum amount of integrated EMG between 20 and 120 msec following probe onset). Trials with no perceptible eye-blink reflex were assigned a score of zero and included in the analysis.

Manipulation check. A manipulation check was done to verify the intended emotion elicitation of each video clip. Because of concerns about the reliability of very young children's self-reports of emotions, we chose instead to code the videotapes of the children's facial and behavioral expressions of affect as they watched each clip. A coding tape was created by randomizing the positive and negative clips for all cases. The clips were dubbed in their entirety and without sound so that the coders were blind to the positive/negative valence of each clip. Each video clip was coded in two epochs: (1) from the start of the clip until the first startle probe and (2) from the first startle probe until the second startle probe. A list of the relevant time stamps for each clip (i.e., start of clip, first startle probe, second startle probe) was provided. Two research assistants trained in coding behavioral expressions of affect independently scored all clips on two dimensions: (1) overall intensity of positive affect (from 0 = no positive affect [no smile or positive verbalization] to 2 = high positive affect [child smiles for most of the epoch; may laugh, giggle, or make positive verbalizations in response to the video]) and (2) overall intensity of fear (from 0 = no fear [child displays no signs of fear or wariness] to 3 = extreme fear [child shows two or more of the following: heavy, labored, or fast-paced breathing, raised eyebrows, wide eyes, "frozen" look, covering eyes]). Intercoder reliability (Cohen's κ) was .74. Where there were discrepancies in codes, the average of the two codes was used. Then, averaged positive and fearful affect scores were computed separately for the two positive clips and the two negative clips. Paired t tests confirmed that the positive clips elicited significantly greater positive affect than the negative clips [$t(27) = 7.23$, $p < .001$] and the negative clips elicited significantly greater fearful affect than the positive clips [$t(27) = -5.68$, $p < .001$].

Data Analysis

A total of 248 probe presentations (8 probes $\times$ 31 children) was administered. Valid EMG results were obtained for 77% of the probes, and valid video scores were obtained for 95% of the probes. Overall, a total of 74% of the probes administered received both EMG scores and video scores. The percentage of valid EMG trials is comparable to those of other research using EMG-measures of eye-blink startle magnitude (e.g., Larson, Ruffalo, Nietert, & Davidson, 2000). The relatively high rejection rate for EMG data is due to its greater sensitivity than video data to excessive prestimulus activity and movement prior to and during the probe. Two primary data analytic strategies were used: (1) within-subjects analyses comparing the EMG scoring and the video scoring at the individual level and (2) group-level analyses of emotion modulation of the startle response.

Within-subjects analyses. To address the relation between the EMG-derived and video-derived measures of eye-blink magnitude, the individual-participant approach to data analysis was considered the most important, since this approach used the data in its raw form without averaging across blink responses. For each of the 31 participants, we examined the correlation between the EMG-derived and the video-derived blink magnitude measures across all probes. Thus, for each individual participant there was a maximum of eight data points (four probes during the positive clips and four probes during the negative clips). The actual number of nonmissing data points available for these within-subjects analyses ranged from three to eight (with a mean of six).

Group-level analyses. Because of large individual differences in the range and distribution of the EMG-derived measure, eye-blink reflex magnitudes were standardized within subjects using the z-transformation of both the EMG and video scores. Group-level analyses focused on the comparison of results using EMG-derived and video-derived measurements. Thus, within-subjects z-transformations were computed using only those probes for which both EMG and video scores were obtained. Then, for each of the 31 participants, we calculated an average standardized EMG-derived blink magnitude in response to probes presented during the positive clips and an average standardized EMG-derived blink magnitude in response to probes presented during the negative clips. Similarly, we calculated an average of the standardized video-derived blink magnitudes. We then examined the correlation between these measures separately for the positive and the negative clips. In addition, we calculated a difference score (blink magnitude during the positive clips subtracted from blink magnitude during the negative clips) separately for the EMG-derived and the video-derived measures. We then examined the correlation between these two difference scores derived from each of the measures. Finally, we investigated whether there was evidence at the group level for valence-dependent emotion modulation of the startle response.

RESULTS

Within-Subjects Analyses

Correlations between EMG and video-scoring methods. For each participant, we first examined the relation between eye-blink magnitude derived from EMG and from video measures across the eight startle probes pre-

sented during the four video clips (two probes per clip). This analysis yielded one correlation for each of the 31 participants. We expected that higher EMG scores would be associated with higher video scores. This being the case, one-tailed tests of significance were used. In Table 1, the Pearson correlations for the EMG and video scores for each child are presented. With the exception of one case in which a negative correlation was observed (Child 4, $r = -.28$), the correlations were all positive and ranged from .52 to 1.0. After the *r*-to-*z* transformations, the average within-subjects correlation across all 31 participants was $r = .84$ (median of raw *r*s = .80). The average correlation for the positive condition was $r = .95$ (median of raw *r*s = .84), and the average correlation for the negative probes was $r = .95$ (median of raw *r*s = .90).

Group-Level Analyses

Group-level correlations between video- and EMG-derived measures of blink magnitude. Using separate within-subjects averages of the blink magnitudes derived from the EMG and video methods, the across-subjects ($N = 31$) correlations between these methods were computed in three ways. First, we examined the correlation of means for probes presented during the positive clips. Second, we examined the correlation of means for probes presented during the negative clips. Third, for each participant we used the negative-minus-positive difference score based on both scoring methods and performed a correlation between these difference scores.

There was a positive correlation between EMG- and video-derived measures on blink magnitude during the positive video clips ($r = .61$, $p < .001$) and the negative video clips ($r = .73$, $p < .001$). The correlation between the two methods for the difference in eye-blink magnitude (mean of negative condition minus mean of positive condition) was also highly significant ($r = .70$, $p < .001$), as can be seen in Figure 1.

For each measurement method, we also examined whether eye-blink magnitude elicited during the negative clips was larger than that elicited during positive clips. Initial analyses showed no significant differences in magnitudes using either method [paired *t* tests: $t(29)_{EMG} = 1.06$, n.s.; $t(29)_{video} = 1.25$, n.s.]. We then considered each video clip separately. Figure 2 shows the box plots, each including the median and interquartile range, of the eye-blink magnitudes obtained in each video clip for

Table 1
Within-Subjects Pearson Correlations for EMG and Video Scores

Child	Within-Subjects Correlation			No. Probes with EMG and Video Scores (Positive/Negative)
	All Probes	Positive Condition	Negative Condition	
1	.85**	.63	.98	3/4
2	.94*	†	.99	1/3
3	.85*	.99	1.00	3/2
4	−.28	†	−.004	1/3
5	.67*	.62	.60	3/4
6	.76	1.00	†	2/1
7	.97**	1.00	.99	2/3
8	.99**	1.00	1.00	2/3
9	.59	.64	.85	3/3
10	.89*	†	.88	1/4
11	1.00**	†	1.00	1/2
12	.90**	.90	.98	4/3
13	.57	.71	.85	4/4
14	.85	†	‡	1/3
15	.84**	.42	1.00	4/3
16	.94*	1.00	1.00	2/2
17	.94*	†	.94	0/4
18	.72*	.99	.52	4/3
19	.60	.68	.74	3/4
20	.59	.98	.24	3/3
21	.87*	†	.89	1/4
22	.74*	.97	.96	3/4
23	.63*	.84	.78	4/4
24	.72*	1.00	.98	2/4
25	.80*	1.00	.87	3/4
26	.83	†	.99	1/3
27	.71**	.70	.98	4/3
28	.79*	.83	.90	3/4
29	.72*	.57	.80	4/4
30	.52	.91	.56	4/4
31	.86**	.65	.85	4/4

†No more than one valid score on both EMG and video. ‡The score on one of the two measures is constant. *$p < .05$, one-tailed. **$p < .01$, one-tailed.

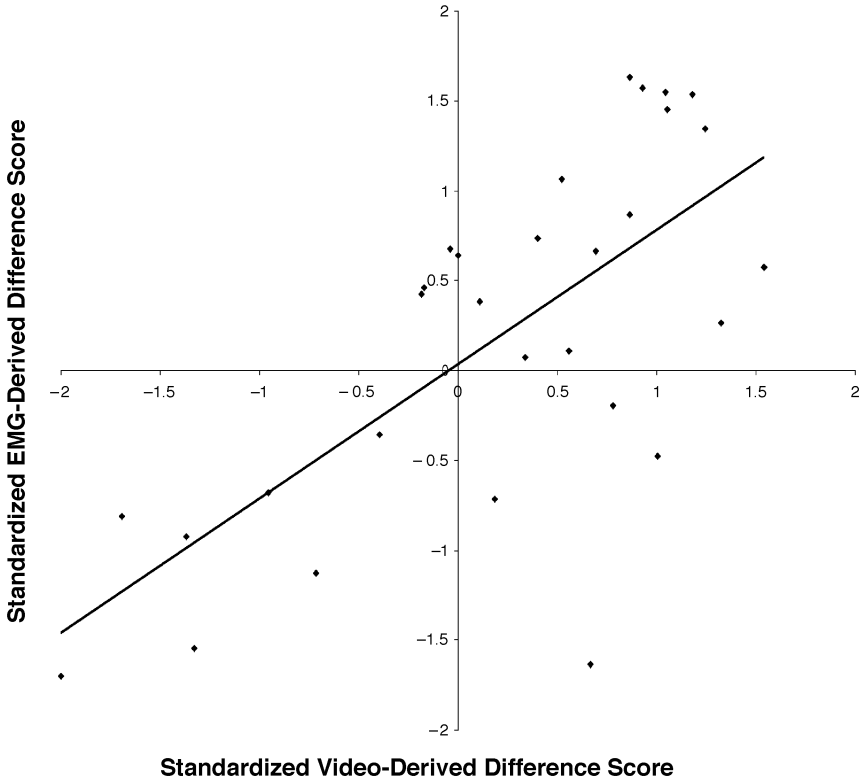

Figure 1. Group-level correlation between EMG- and video-derived measures of standardized difference scores in startle magnitude (negative condition − positive condition).

each method. Results of paired *t* tests showed that there was no significant difference in the means of the two positive clips using either the EMG [$t(21) = -1.89$, n.s.] or the video-derived [$t(21) = -1.90$, n.s.] method. However, there was a significant difference in the means of the two negative clips using either the EMG [$t(28) = -3.32$, $p = .003$] or the video-derived [$t(28) = -4.22$, $p < .001$] method. Therefore, for each method, we used the eye-blink magnitude averaged across the positive clips but analyzed the negative clips separately. Results of paired *t* tests showed that the mean of the standardized eye-blink magnitude measured during the first negative clip did not differ significantly from the mean magnitude measured during the positive clips using either the EMG [$t(28) = -0.44$, n.s.] or the video-derived [$t(28) = -0.71$, n.s.] method. The startle responses to the probes in the second negative clip, however, were significantly larger than those in the positive clips for both the EMG [$t(28) = 2.17$, $p = .038$] and the video-derived [$t(28) = 2.80$, $p = .009$] methods.

DISCUSSION

The purpose of this study was to examine the reliability and validity of a method of assessing the magnitude of emotion-modulated startle outside of a laboratory setting on the basis of video scoring of children's eye blinks in response to acoustic startle probes presented during emotion-eliciting video clips. The results indicate that reflex magnitude in response to brief acoustic startle probes can be reliably coded from a video record. The interrater reliability of this coding was high. Moreover, the magnitude of the startle reaction assessed with video coding was correlated with simultaneously obtained EMG measures of blink magnitude, both within participants and across the group. The within-participants correlations were, with a single exception, moderate to high. The single exception ($r = -.28$, n.s.) was from a child who had exceptionally low EMG values, with the largest blink below 6 μV. The group-level correlations were consistently high.

There were a number of cases ($n = 9$) for which the within-subjects point-to-point correlation coefficients were moderate but not significant. It is not surprising that the correlation between a continuous measure such as the EMG and a categorical variable such as the video code shows some attenuation. Importantly, however, there is not a systematic relationship between number of valid trials and within-subjects correlation. Five cases had six to eight valid responses, for which the correlations ranged from .70 to .75. Another six cases with seven or eight valid probes ranged from .79 to .90. It is likely that the EMG measure reflects both visible and nonvisible components of muscle tension. Furthermore, as will be

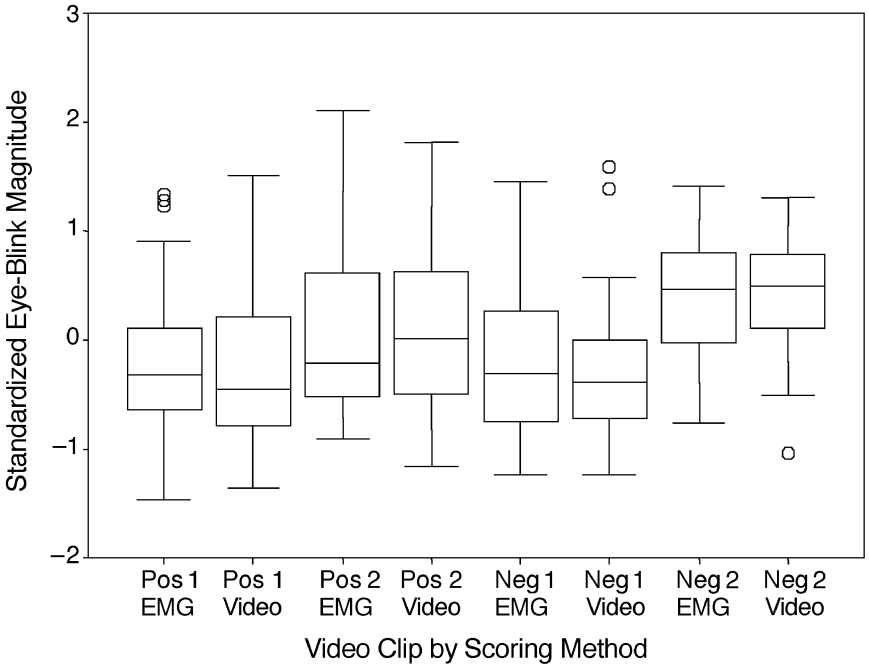

Figure 2. Box plots of standardized scores (median and interquartile range) for each video clip and each method.

discussed in greater detail below, the two measures have different temporal parameters. Thus, the fact that the correlations show some range and are not perfect is expectable.

The analysis of both the video and the EMG data partially confirmed the valence-dependent emotion modulation hypothesis. That is, when the emotion-modulation hypothesis was examined in a comparison between the averaged eye-blink magnitudes during the two positive clips and those during the two negative clips, we found no differences. However, when only the second negative clip and both positive clips were considered, the expected pattern of emotion modulation effects was clearly evident, both in the EMG data and in the video-scored results. As was predicted by the emotion-modulation hypothesis, the probes inserted on the second negative clip produced startle reactions significantly larger in magnitude than those produced by the probes presented against a positive foreground video.

Without a doubt, the temporal resolution of the EMG measure and that of the video measure differ substantially, raising questions as to how well they captured the same process. The EMG-derived startle response peak magnitude occurs between 20 and 120 msec after onset. The first, fastest, and most stable element in the sequence of the visible startle reaction is the sudden contraction of the orbicularis oculi (seen as movement of the eyelid), which typically begins 30–50 msec after the presentation of an acoustic probe (Lang et al., 1990). In comparison, given the resolution of the 8-mm tape (33.3 msec/frame), the peak of the eye-blink response is likely to occur within the first three frames after onset, suggesting that the visible eye-blink response falls within the temporal limits imposed by the 8-mm video recording method used. Importantly, however, because the EMG-measured peak of startle can have a duration of one or a few milliseconds and we do not know the time course of the video-rated blink, it is possible that a peak occurring within the first three frames after onset would not be captured on the tape. Nonetheless, the strength of the within-subjects correlations indicates that what is being measured from the video coding of the eye blink corresponds reasonably well with the process reflected in the EMG measure, despite the superior temporal resolution of the latter.

Empirically, we pursued this issue further in two ways, each suggesting that the process captured by the video coding is comparable to that reflected in the EMG measure. First, we reviewed the scoring sheets from the video coding, which contained frame-by-frame records of the time of each probe and the onset and completion of blinks. The eye blinks appeared in one to four frames, with the smallest blinks seen in a single frame and the larger blinks occupying up to four frames. Furthermore, given the 33-msec sampling of the camera, the blink seen on the video could never be more than 16 msec from the peak. These results are comparable with the temporal pattern obtained with the EMG measure, as was described above. Second, we took 3 participants and down-sampled the EMG to a resolution comparable to that of the video. We then computed the correlation between the startle magnitude on the basis of this down-sampled

EMG and the original high-resolution EMG. The correlation was .76 ($p < .001$), again suggesting that the two methods captured substantially the same process. This indicates that the startle response magnitude derived from each of these measures will be substantially similar. Related issues were the greater number of codable trials obtained with the video-coding method and the concern that these could be false positives. A review of the probes that received only video scores showed that there was a larger number of valid video than EMG probes because the video technique is less prone to corruption due to certain kinds of artifacts. For example, tension around the eyes will activate the orbicularis oculi, making the EMG signal difficult to analyze but not interfering with the video coding. Similarly, gross body movement can corrupt the EMG, whereas the video method can still extract the blink. Thus, it appears that the larger number of video-scored probes does not reflect false positives with the video method; the video method might even be advantageous in this regard. However, it is also important to note that the video method may not be able to exclude nonstartle blinks if they occur around the time of stimulus onset.

Several practical issues and limitations to the video-based approach of assessing the magnitude of the startle response remain to be addressed and refined. An important area of future work involves the development and use of stimulus tapes. It would have been informative to know more about the actual affect-inducing qualities of the video clips used in this study. Although a manipulation check verified that the intended emotions were being elicited in the positive and negative clips, the results suggest that the first negative clip was not sufficiently fear inducing to elicit increased eye-blink magnitude in comparison with that elicited by the positive clips. More specifically, (1) because we did not ask the children to rate the valence of the clips, we do not know the extent to which the children themselves experienced the clips as positive or negative; (2) we know that some of the children had seen some of the full-length videos previously, but we did not systematically ask them about their familiarity with the clips; (3) there was no "neutral" condition, which limits the evaluation of affect modulation to the comparison of positive versus negative effects rather than extending it to the potentially more informative comparison of positive versus neutral and negative versus neutral effects; and (4) the extent to which the positive clips could elicit positive affect may have been affected by the fact that they were presented first. Other work (see the discussion in Davidson & Irwin, 1999) has shown that the first trials of many physiological protocols may not reflect even a neutral or baseline condition very well, but rather a more negative condition owing simply to the novelty and anxiety of the experience of being hooked up to physiological monitoring devices. In the present study, this may have resulted in a positive condition confounded by order of presentation, attenuating the difference between the positive and negative conditions. Habituation effects can also affect startle reactivity (see Sutton et al., 1997). In this study, habituation would work against the hypothesis of greater magnitude of startle during the fear-inducing clips than during the happiness-inducing clips owing to the order of presentation of the videos (positive first, then negative). The fact that we did observe greater startle magnitude during one of the fear-inducing clips in comparison with the happiness-inducing clips despite the fact that the fear-inducing clips always occurred last suggests that the actual difference in startle magnitude between these conditions may have been larger than was observed, but was somewhat masked by habituation. These potential confounds due to order of presentation should be considered carefully in the design of future studies. Finally, it should be emphasized that this study shows that the video method is appropriate for measuring startle magnitude in situations outside of a laboratory setting. However, the video method does not provide measures of other startle parameters such as latency and duration.

In the process of developing a visual coding system that could be used to reliably classify a continuous physiological response into discrete categories, several issues arose that resulted in the establishment of some important coding guidelines. Between the prototypic examples of the response categories (e.g., "perfect quarter-blink" and "perfect half-blink"), there are fuzzy areas at the boundaries of the categories (e.g., one person's "very strong 3" may be another person's "very weak 4"). As such, to facilitate intercoder reliability, the following guideline was established at the outset: When a score was judged by a coder to be extremely borderline between two categories, rounding up was specified. Other practical guidelines were established as well. For example, blinks in response to the startle probe were not coded if the probe occurred while a natural blink was ending or had already begun; the same eye *must* be coded by both coders (e.g., the right or the left eye, whichever is on the camera side, to optimize coding; in this study, it was critical to code the eye under which the spot electrodes were placed). Owing to the nature of the coding system as well as to the variability between individuals in a number of physical characteristics that can make coding more complicated (e.g., eye shape, eyelid structure), it is important that coders be trained using an objective training tape containing the full range of responses and a number of different individuals. Alternatively, with improving computer technology, it is now possible to increase the reliability of the video method by manually digitizing the location of the top and bottom edges of the eyelids and either using off-the-shelf software or constructing a point-and-click digitizer for frame-by-frame analysis of the video data. It may also be possible to automate the measurement of frame-to-frame changes in the distance between the top and bottom eyelids.

In summary, these data demonstrate that a video-based coding methodology is a reliable and valid proxy for EMG-based measures of the magnitude of the startle

response for studies of children conducted outside of the laboratory. It is important to note, however, that very small responses may occasionally not be captured with the video-based method, making it difficult to assess startle response probability. This method also may not be able to exclude nonstartle blinks if they occur around the time of stimulus onset. Finally, it cannot be used to assess the latency or duration of the startle response. Nevertheless, the use of this video-based assessment of emotion-modulated startle has the potential to expand the repertoire of methods available to researchers interested in obtaining information on emotional reactivity in settings in which traditional physiological measurement is not possible.

REFERENCES

Balaban, M. T. (1995). Affective influences on startle in five-month-old infants: Reactions to facial expressions of emotions. *Child Development*, **66**, 28-36.

Cook, E. W., Hawk, L. W., Davis, T. L., & Stevenson, V. E. (1991). Affective individual differences and startle reflex modulation. *Journal of Abnormal Psychology*, **100**, 5-13.

Davidson, R. J. (1992). Emotion and affective style: Hemispheric substrates. *Psychological Science*, **3**, 39-43.

Davidson, R. J., Ekman, P., Saron, C. D., Senulis, J. A., & Friesen, W. V. (1990). Approach–withdrawal and cerebral asymmetry: Emotional expression and brain physiology. I. *Journal of Personality & Social Psychology*, **58**, 330-341.

Davidson, R. J., & Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Science*, **3**, 11-21.

Davidson, R. J., & Sutton, S. K. (1995). Affective neuroscience: The emergence of a discipline. *Current Opinion in Neurobiology*, **5**, 217-224.

Davis, M. (1992). A neural systems approach to the study of the amygdala, fear and anxiety. In J. M. Elliott, D. J. Heal, & C. A. Marsden (Eds.), *Experimental approaches to anxiety and depression* (pp. 45-71). Chichester, U.K.: Wiley.

Flaten, M. A. (1993). A comparison of electromyographic and photoelectric techniques in the study of the classical eyeblink conditioning and startle reflex modification. *Journal of Psychophysiology*, **7**, 230-237.

Goldsmith, H. H. (1993). Temperament: Variability in developing emotion systems. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 353-364). New York: Guilford.

Grillon, C., Dierker, L., & Merikangas, K. R. (1997). Startle modulation in children at risk for anxiety disorders and/or alcoholism. *Journal of the American Academy of Child & Adolescent Psychiatry*, **36**, 925-932.

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, **50**, 372-385.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, **97**, 377-395.

Larson, C. L., Ruffalo, D., Nietert, J. Y., & Davidson, R. J. (2000). Temporal stability of the emotion-modulated startle response. *Psychophysiology*, **37**, 92-101.

Patrick, C. J., Bradley, M. M., & Lang, P. J. (1993). Emotion in the criminal psychopath: Startle reflex modulation. *Journal of Abnormal Psychology*, **102**, 82-92.

Sutton, S. K., Davidson, R. J., Donzella, B., Irwin, W., & Dottl, D. A. (1997). Manipulating affective state using extended picture presentations. *Psychophysiology*, **34**, 217-226.

Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response: A new measure of emotion? *Journal of Abnormal Psychology*, **97**, 487-491.

NOTE

1. A progressive-scan (noninterlaced) camcorder that captures 30 frames/sec has the distinct advantage of minimizing movement artifact. However, there is also an advantage in using an interlaced scan that could capture 60 frames/sec, at the cost of some vertical spatial resolution, and later de-interlacing the video using currently available software. Doubling the temporal resolution would be especially advantageous for coding blinks that are captured in just one frame. Moreover, by turning the camera sideways, the loss of spatial resolution due to interlacing would be horizontal rather than vertical, which is crucial for the detection of blinks.

(Manuscript received March 27, 2000;
revision accepted for publication March 18, 2003.)