Developmental Psychology
1999, Vol. 35, No. 2, 505–513

0012-1649/99/$3.00

An Event-Based Analysis of the Coordination of Early Infant Vocalizations and Facial Actions

Marygrace E. Yale, Daniel S. Messinger, Alan B. Cobo-Lewis, D. Kimbrough Oller, and Rebecca E. Eilers
University of Miami

This study used an event-based approach to provide empirical evidence regarding the nature of coordination in 3- and 6-month-old infants. Vocalizations and facial actions of 12 normally developing infants interacting with their caregivers were coded. Coded vocalizations and facial actions were considered coordinated when they temporally overlapped. Results indicate that infants coordinated their vocalizations and facial actions more than expected by chance. Coordinated events were governed by 2 sequence patterns. When 2 communicative events were temporally associated across modalities, 1 event tended to be completely embedded within the other, and vocalizations tended to end before facial actions. This study provides new information about how infant communication is structured, confirms results from other coordination studies, and describes a new method for analysis of event-based data.

Infants communicate through various modalities, including facial expressions, vocalizations, and gaze direction. Most studies on the coordination of actions in these domains have focused on the co-occurrence of infant gaze direction with facial expressions or vocalizations (Fullmer & Messinger, 1997[1]; Kaye & Fogel, 1980; van Beek, Hopkins, & Hoeksma, 1994; Weinberg & Tronick, 1994). Surprisingly, little is known about whether infants create coordinated patterns of vocal and facial signals at levels greater than one would expect by chance. In the present study, we addressed this question by using an event-based analysis that was capable of examining the sequential makeup of coordinated events involving facial expressions and vocalizations.

The majority of coordination studies have focused on the way that early communicative behaviors are coordinated with infant gaze direction during face-to-face interactions (Fullmer & Messinger, 1997; Kaye & Fogel, 1980; Messinger, 1994; van Beek et al., 1994; Weinberg & Tronick, 1994). Results have indicated that the amount of smiling and vocalizing increases when infants gaze at their mothers during the first 6 months of life (Fullmer & Messinger, 1997; Kaye & Fogel, 1980; Messinger, 1994; van Beek et al., 1994). These results suggest the importance of visually mediated social contact to infant emotional signaling in the facial and vocal domains.

Yet, few studies have examined the coordination of vocalizations and facial expressions, two communicative modalities that index affective states (Austin, 1962; Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Demos, 1982; Ekman, 1972; Tomkins, 1963). Weinberg and Tronick (1994) examined the coordination of various facial expressions thought to indicate specific emotions with several communicative behaviors, including vocalization, gaze, posture, and gesture. Four distinct affective configurations were identified. These configurations involved the co-occurrence of facial displays of joy, interest, sadness, and anger with different vocalization types, gestures, postures, and direction of gaze (Weinberg & Tronick, 1994). In a similar vein, Fullmer and Messinger (1997) found that fussy vocalizations tended to co-occur with negative facial expressions and that noncomplaint vocalizations tended to co-occur with positive facial expressions.

Two alternative perspectives on coordination may be termed *time-* and *event-based approaches* (see Figure 1). Previous studies of coordination (Fullmer & Messinger, 1997; Kaye & Fogel, 1980; Weinberg & Tronick, 1994) used a time-based approach in which the unit of coding and analysis was a fixed time interval (e.g., 1 s or one video frame). Time-based analysis rests on the assumption that observed coordination is indexed by the total duration of intervals in which the two events of interest co-occur. However, in an event-based approach, the focus is on the temporal sequencing of whole actions from different communicative modalities that have some temporal overlap. For example, if a smile overlaps a vocalization, it may be reasonable to assume that the two events

Marygrace E. Yale, Daniel S. Messinger, Alan B. Cobo-Lewis, D. Kimbrough Oller, and Rebecca E. Eilers, Department of Psychology, University of Miami.

Alan B. Cobo-Lewis is now at the Department of Interdisciplinary Studies, University of Maine. D. Kimbrough Oller is now at the Department of Communication Sciences and Disorders, University of Maine. Rebecca E. Eilers is now at the College of Arts and Sciences, University of Maine.

Portions of these data were presented at the biennial meeting of the Society for Research in Child Development, Washington, DC, April 1997.

This research was supported by Grant R01 DC00484 from the National Institute of Deafness and Other Communication Disorders.

We thank the families who participated in this longitudinal study and the research assistants who coded the data.

Correspondence concerning this article should be addressed to Marygrace E. Yale, Department of Psychology, University of Miami, P.O. Box 249229, Coral Gables, Florida 33124-0721. Electronic mail may be sent to myale@umiami.ir.miami.edu.

[1] Both the present study and that of Fullmer and Messinger (1997) were products of the same broad longitudinal research program. However, the two efforts were based on different analysis approaches, and only parts of some of the same recordings were used. Some portions of the 6-month recordings of 11 of the 12 infants observed in the present study were also used by Fullmer and Messinger.

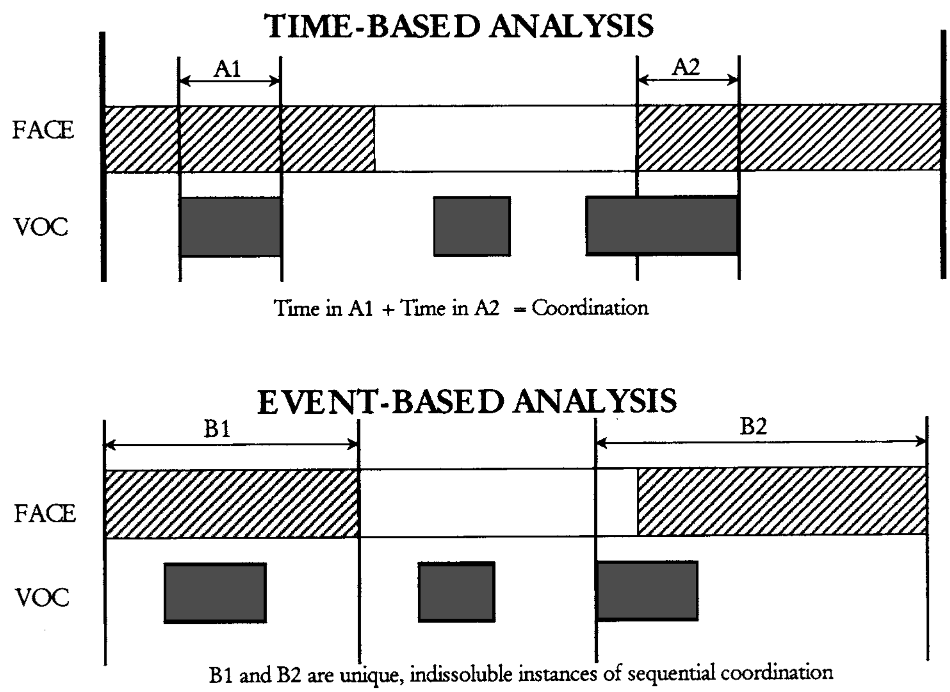

Figure 1. Contrast between time-based and event-based approaches to the study of coordination. VOC = vocalization; FACE = facial action.

create a coordinated action, even though the two events do not overlap entirely. An event-based approach suggests that when two behaviors from different modalities temporally overlap, they create a communicative signal whose coordination may be best understood by examining the temporal sequences of the constituent actions. We submit that it is important to understand not only that communicative events go together (as can be indicated by a time-based approach) but also how infants temporally coordinate these events (as can be indicated by an event-based approach).

Few studies have considered the coordination of communicative behaviors from an event-based perspective, in part because this perspective requires new quantitative methods for the analysis of data. In this study, we used a computer-based coding and analysis system that identified sequences of observed events involving coordination between two communicative modalities. Then we used a new statistical approach (Cobo-Lewis, 1998) to determine the degree of event-based coordination that could be expected by chance. By comparing observed versus expected levels of coordination, we were able to determine if observed levels of coordination were greater than those expected by chance. To our knowledge, this represents the first event-based statistical test of cross-modal coordinated action in infancy.

Purpose of the Present Study

This study focuses on the coordination of infant facial actions and early vocalizations from an event-based perspective. We examined the specific temporal relationships between communicative modalities in these coordinated events among infants in the first 6 months of life. We also considered the effects of different interactive contexts and of age on the extent of observed coordination. In addition, we analyzed the components of these coordinated events, specifically the temporal relationships between communicative modalities and the affective tone of facial expressions involved.

Method

Participants

Of 25 infants recruited for a larger longitudinal study, 12 met the present study's selection criterion of producing at least 10 vocalizations during visits when the infants were 3 and 6 months old. All participants were recruited by mail solicitation on the basis of Florida state health department birth records and were from middle- to upper-middle-class families as ascertained by the mothers' educational levels and zip codes. Recruited participants were all healthy full-term infants with unremarkable pre- and postnatal medical histories. Eight of the 12 participants were White non-Hispanic, and 4 were White Hispanic. The ethnic backgrounds of the participants in the present study did not differ from the backgrounds of the participants in the larger study. Mean ages of the 12 participants (6 boys and 6 girls) at the 3-month and 6-month visits were 11 weeks 6 days and 26 weeks 5 days, respectively.

Apparatus

All visits were recorded in a sound-attenuated chamber (11 ft [3.35 m] × 11 ft [3.35 m] × 6.5 ft [1.98 m]) with high-fidelity audio recording from a microphone placed above the participants' heads. During both the 3- and 6-month visits, the caregiver sat in a chair facing the infant. The infant was placed in an infant seat mounted on a table during the 3-month visit and in a high chair during the 6-month visit. The views from two cameras, one offering a full frontal view of the infant's face and one offering a three-quarter frontal view of the caregiver's face, were combined in split-screen format and recorded on a JVC Super VHS video recorder (Model BR-S800-U).

Procedure

During each visit, an experimenter sat in a chair, behind the caregiver and out of the infant's view. As each new episode of the session was initiated, the tester provided instructions to the parent. The session began with a 3-min face-to-face play segment. The experimenter instructed the caregiver, "Play with your infant as you normally would do at home. Talk to him/her and try and get your baby to talk back to you." A modified still-face episode began with instructions to the caregiver to sit back and look at a picture placed on the wall behind the infant. The caregiver was told to stare at the picture and maintain a still face. The caregiver was informed not to respond to the infant (i.e., not to look down, smile, or nod at the infant) until otherwise instructed. This modification of the traditional still-face paradigm (Tronick, Als, Adamson, Wise, & Brazelton, 1978) was intended to provide opportunities for the infant to make communicative bids to the unresponsive caregiver while minimizing fussiness (in a traditional still face, the mother stares directly at the baby). The modified still-face episode lasted 1 min and was followed by another 3-min face-to-face play episode in which the caregiver was instructed to cease the still face and to play with the infant as he or she had at the beginning of the session. During the visits, 10 of the participants interacted with their mothers, and 2 interacted with their fathers.

Portions of the session during which instructions were given to the caregiver and during which the infant or caregiver shifted position to achieve improved camera angles were not analyzed. Of the 12 participants, 5 did not have a second face-to-face play episode immediately following the modified still face at their 3-month visit. For these participants, we substituted another 3-min face-to-face interaction from the end of the session. We conducted a t test to compare the proportion of time spent vocalizing in the substituted face-to-face interaction episode and the proportion of time spent vocalizing in the second face-to-face play episode for the 5 infants in question and for the remaining 7 infants at the 3-month visit. The two groups did not differ in the proportion of time spent vocalizing, $t(10) = -0.497, p > .50$, nor in the proportion of time spent in coded facial actions, $t(10) = -0.755, p > .45$.

Coding

Coding was performed using Action Analysis Coding and Training (AACT, 1996; Oller, Yale, & Delgado, 1997). The AACT system provides computer-assisted observational coding, with the target behaviors entered into a computer directly controlling the Super VHS videotape machine with single-frame accuracy and automatic time-code capture. An analysis module included in AACT allows for graphical display of temporal relationships across different dimensions of action.

Facial action coding. Facial coding involved combinations of facial action units from Ekman and Friesen's (1978) Facial Action Coding System (FACS) and Oster and Rosenstein's (in press) Baby FACS. This system allowed us to analyze facial expressions from an anatomically based perspective. Action units identified by the FACS systems were combined into more general measures of facial expression (Camras, Oster, Campos, Miyake, & Bradshaw, 1992). Coded events included smiles (lip-corner raises) and anger–distress–sadness displays. The latter entailed lowered brows with (a) a prototypical open, squared anger mouth; (b) tightened lips and sideways lip stretch; or (c) a chin raise. Both smiles and anger–distress–sadness displays were considered coded facial actions and were distinguished from neutral facial actions (the absence of these facial codes). Infant facial actions were coded in slow motion (one sixth of regular tape speed), with onset and offset times of each facial action unit recorded by two coders, each trained by a FACS-certified coder.

Vocalization coding. The infants' vocalizations were coded at the utterance level by two coders trained in identifying and classifying different types of early infant vocalizations. Vegetative vocalizations (i.e., burps, hiccups, coughs, and sneezes) were not included in subsequent analyses (Oller, 1980; Oller & Eilers, 1988). Coders identified each vocalization in real time and then localized onset and offset times more precisely.

Temporal associations between vocalization and facial action events. Different vocalization–facial action sequence patterns were determined by using AACT software (Oller et al., 1997). A vocalization event and a facial action event could be temporally associated in any of the following ways (see Figure 2):

A. *Noncoordinated events:* These events involved no temporal overlap between a vocalization and a coded facial action:

1. *Vocalization alone:* a vocalization that occurred in the absence of any coded facial action.
2. *Facial action alone:* a coded facial action that occurred in the absence of any vocalization.

B. *Coordinated events:* These events involved a temporal overlap between a vocalization and a facial action and were further categorized as follows:

1. *Vocalization in facial action:* a vocalization that began and ended within a coded facial action.
2. *Facial action in vocalization:* a coded facial action that began and ended within a vocalization.
3. *Vocalization after facial action:* a vocalization that began within but ended after a coded facial action.
4. *Vocalization before facial action:* a vocalization that began before but ended within a coded facial action.

Reliability. Interobserver reliability for coding of facial actions and vocalizations was assessed for 20% of the sessions. The average Cohen's (1960) kappa for facial actions was .63, with an average agreement of 74% (individual kappas and average agreement for smiles were .70 and 90%, respectively, and for anger–distress–sadness displays were .69 and 91%, respectively). For vocalizations, the average Cohen's kappa was .73, with an average agreement of 96%. The primary theoretical and analytic focus of this research was the coordination of vocalizations and facial actions. To ascertain the reliability of this dependent measure, a coder viewed the sessions in real time and coded how each vocalization was temporally associated with the infant's facial actions. This coder agreed with the associations (made through the AACT system's analysis module) to the level of the specific association type (facial action in vocalization, vocalization in facial action, etc.) 87% of the time.

Bootstrapping Procedure

Having measured the observed frequencies of different types of coordinated sequences, we then used a bootstrapping procedure (e.g., Efron & Tibshirani, 1993; Mooney & Duval, 1993) to determine the frequencies expected by chance. The bootstrapping program (Cobo-Lewis, 1998) used the actual frequencies, durations, and sequential transition probabilities of vocalizations and facial actions separately in each session as raw data. It then constructed 1,999 simulated sessions for each child, at each age, and for each protocol episode. In each simulated session, behaviors in the vocal dimension and in the facial action dimension were simulated independently of the other. That is, temporal associations between vocalization and facial action events were computed from each simulated session to calculate the estimated sampling distribution for each type of temporal association under the null hypothesis that the two dimensions were independent. Thus, the procedure computed the frequency of each type of coordinated event expected by chance. The frequencies of temporal associations between vocalization and facial action events observed in the actual sampled data were compared with frequencies from the simulated distributions to determine if the vocalizations and facial actions were coordinated more than one would expect by chance for each protocol episode, at each age, for each infant. A total of 143,928 computer-simulated sessions were modeled for the present study.

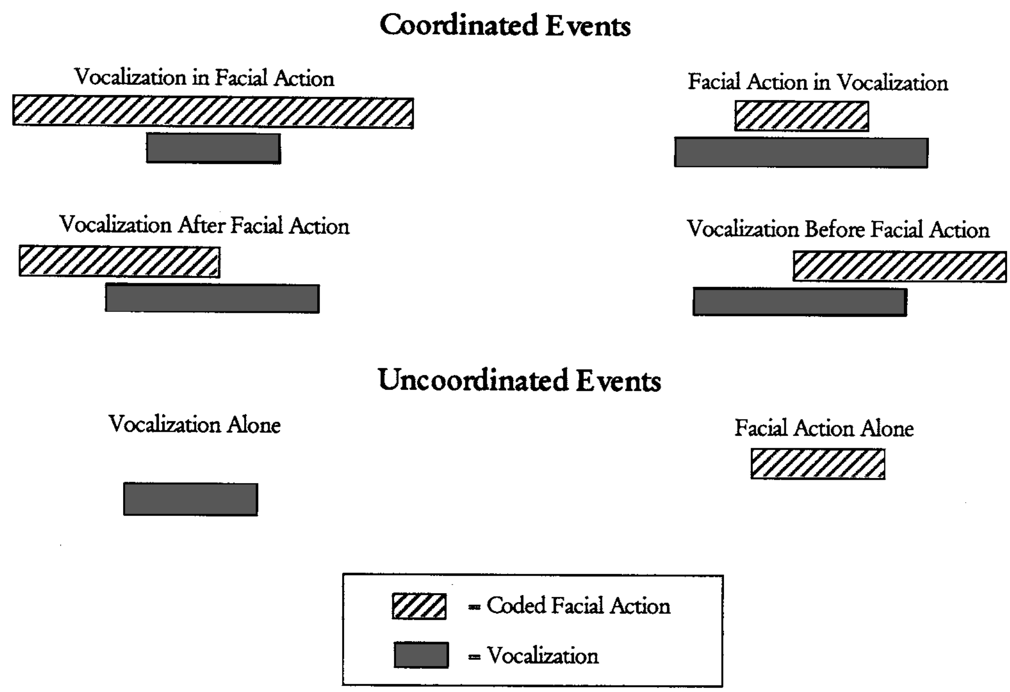

Figure 2. Temporal associations between vocalization and facial action events.

The bootstrapping procedure yielded z scores that quantified the temporal association between vocalization and facial action events observed in the data relative to the temporal association expected by chance. Each z score was computed as the number of observed events minus the mean number of events expected by chance, divided by the standard deviation of the sampling distribution expected by chance. A positive z score indicates that the temporal association between vocalization and facial action events in question occurred more than expected by chance. A negative z score indicates fewer occurrences than expected by chance.

Results

Descriptive Results of Facial Actions and Vocalizations

Table 1 presents observed and expected frequencies and durations of facial action and vocalization events. These data summarize characteristics of each communicative modality considered alone and provide information with which to assess the adequacy of the bootstrapping procedure. At both 3 and 6 months, facial actions occupied about twice as much time as vocalizations and had average durations that were twice as long. In Table 1, the adequacy of the bootstrapping procedure is exhibited by agreement between expected and observed numbers.

To summarize characteristics of cross-modal coordination, Table 2 presents summary statistics for the temporal associations between vocalization and facial action events observed in the actual data and those expected from the simulated data. In Table 2, systematic coordination between modalities is exhibited by disagreement between expected and observed numbers.

Analysis of Coordinated Events

We conducted a 2 (age) × 3 (episode) × 2 (tone) repeated measures analysis of variance (ANOVA) to test the hypothesis that infants would coordinate vocalizations and coded facial actions more than expected by chance.[2] This inclusion of tone in the analysis allowed us to examine whether vocalizations were differentially coordinated with smiles (positive tone) or anger–distress–sadness displays (negative tone). In this and all subsequent analyses involving variables with more than two levels, Greenhouse–Geisser (Greenhouse & Geisser, 1959) corrections were used. The grand mean of the z scores computed for the coordinated events ($M = 0.884$, $SD = 0.093$) differed significantly from zero, $F(1, 11) = 91.94$, $p < .001$, indicating that the number of observed coordinated events was significantly greater than expected by chance.

[2] The bootstrapping procedure may have slightly overestimated the frequency of communicative events, particularly facial actions, and slightly underestimated the mean duration of communicative events, particularly vocalizations, in the expected data (see Table 1). The overestimation of the number of facial actions may have led to an artificial deflation of all z scores because an increase in the number of facial actions increased the expected likelihood of coordinated events, thus reducing the difference between observed and expected levels. Shorter durations in the expected data, however, may have increased the likelihood of a coordinated event being classified as vocalization in facial action because shorter vocalizations would be more easily embedded in facial actions. This artificial increase in the expected number of vocalization-in-facial-action types would have artificially deflated the sequence type with the highest mean z score. However, the shorter durations of vocalizations may have decreased the likelihood of facial-action-in-vocalization, vocalization-before-facial-action, and vocalization-after-facial-action classifications, thus artificially inflating the mean z scores of the lowest three types. It is not clear to what extent the small biases in the bootstrapping procedure may have influenced the results. Refinements to the bootstrapping estimations are being made for the next round of studies. In the meantime, the outcomes reported in the present study have been interpreted conservatively.

Table 1
Mean Frequencies (Per Minute) and Mean Durations (in Seconds) of Observed and Expected Estimates of Coded Behaviors

Behavior	At 3 months		At 6 months	
	Observed	Expected	Observed	Expected
		Face-to-face play		
Facial actions				
Frequency	4.22 (0.80)	4.43 (0.79)	2.93 (0.36)	3.16 (0.37)
Duration	1.66 (0.15)	1.64 (0.15)	1.14 (0.20)	1.40 (0.19)
Vocalizations				
Frequency	3.99 (0.85)	4.22 (0.91)	1.92 (0.30)	1.79 (0.21)
Duration	0.65 (0.10)	0.62 (0.09)	0.60 (0.07)	0.57 (0.06)
		Modified still face		
Facial actions				
Frequency	2.44 (0.64)	2.63 (0.64)	3.05 (0.45)	3.46 (0.52)
Duration	0.58 (0.11)	0.55 (0.10)	1.69 (0.52)	1.66 (0.52)
Vocalizations				
Frequency	3.56 (0.74)	3.71 (0.79)	6.52 (1.14)	6.86 (1.16)
Duration	0.50 (0.09)	0.44 (0.08)	0.77 (0.14)	0.68 (0.11)
		Reunion face-to-face play		
Facial actions				
Frequency	3.91 (0.51)	4.08 (0.50)	3.66 (0.49)	3.82 (0.48)
Duration	1.90 (0.46)	1.84 (0.46)	1.94 (0.22)	1.96 (0.23)
Vocalizations				
Frequency	4.21 (0.63)	4.34 (0.64)	5.26 (0.94)	5.38 (0.87)
Duration	0.67 (0.08)	0.64 (0.08)	0.66 (0.08)	0.64 (0.07)

Note. Standard deviations are in parentheses.

There was a significant main effect for tone, $F(1, 11) = 35.49, p < .001$, indicating that of the two types of coded facial actions, anger–distress–sadness expressions (mean z score = 0.859) were more likely than smiles (mean z score = 0.405) to be coordinated with vocalizations. Neither the main effects for age or episode nor any of the two- or three-way interactions were significant. In addition, analyses of the noncoordinated events (vocalization-alone and facial-action-alone types) indicated that these events occurred less than expected by chance: for vocalization alone, $F(1, 11) = 29.54, p < .001$, and for facial action alone, $F(1, 11) = 115.58, p < .001$.

Analysis of the Specific Sequencing Types of Coordinated Events

An event-based approach allowed for examination of the temporal sequencing of vocalizations and facial actions. We conducted a 2 (age) × 3 (episode) × 4 (type) repeated measures ANOVA to examine the z scores computed for the specific sequence types mentioned above from the observed data and the simulated expected data. There was a significant main effect for type, $F(2.36, 25.95) = 3.92, p = .027$, indicating that certain coordinated sequences were more likely than others. Figure 3 presents the mean z scores of the four sequence types. Neither the main effects for age or episode nor any of the two- or three-way interactions were significant. A Tukey's honestly significant difference procedure conducted on the four sequence types revealed that the mean of the z scores for the vocalization-after-facial-action sequence type was significantly smaller than the vocalization-in-facial-action sequence type ($p < .05$). No other pairwise comparisons were significant.

How Coordinated Events Were Sequenced

The four sequence types were then collapsed into categories that reflected theoretical similarities (see Figure 3). In two of the sequence types (vocalization in facial action and facial action in vocalization), one communicative event was temporally embedded within the other, whereas this was not the case in the other two sequence types (vocalization after facial action and vocalization before facial action). In addition, the vocalization-in-facial-action and vocalization-before-facial-action sequence types were similar in that the vocalization ended before the coded facial action; this distinguished them from the facial-action-in-vocalization and vocalization-after-facial-action sequence types, in which the coded facial action ended before the vocalization. We conducted bootstrap analyses to calculate z scores for embedded versus nonembedded types, for vocalization-end-first versus facial-action-end-first types, and for the interaction between the two variables. Repeated measures ANOVAs—2 (age) × 3 (episode) × 2 (embed) and 2 (age) × 3 (episode) × 2 (end first)—indicated that the sequence types in which a communicative event was temporally embedded within another were more favored than sequence types in which there was no embedding, $F(1, 11) = 6.15, p = .031$, and that the sequence types in which the vocalization ended before the facial action were more favored than sequence types in which the

Table 2
Summary of Mean Numbers of Temporal Associations Between Vocalization and Facial Action Events

Sequence type	At 3 months: Observed	At 3 months: Expected	At 6 months: Observed	At 6 months: Expected
Face-to-face play				
Vocalization and smile	10.25 (9.43)	8.59 (8.66)	4.00 (3.44)	2.57 (1.90)
Vocalization and frown	4.75 (7.88)	2.78 (4.96)	1.59 (2.54)	0.41 (0.67)
Vocalization in facial action	8.08 (6.13)	5.63 (4.24)	2.50 (2.35)	1.61 (1.20)
Facial action in vocalization	2.08 (4.81)	0.89 (1.93)	0.75 (1.54)	0.19 (0.23)
Vocalization before facial action	3.33 (4.70)	2.42 (3.76)	1.58 (1.68)	0.58 (0.42)
Vocalization after facial action	1.50 (2.75)	2.43 (3.76)	0.75 (0.62)	0.59 (0.42)
Modified still face				
Vocalization and smile	1.00 (2.37)	0.54 (0.96)	3.08 (2.81)	2.40 (2.58)
Vocalization and frown	1.50 (2.91)	0.65 (1.00)	2.17 (3.24)	1.75 (2.46)
Vocalization in facial action	0.67 (1.37)	0.37 (0.60)	2.08 (2.47)	1.69 (1.89)
Facial action in vocalization	0.83 (1.19)	0.23 (0.37)	0.83 (1.75)	0.47 (0.88)
Vocalization before facial action	0.58 (1.38)	0.30 (0.37)	1.25 (1.29)	0.97 (0.88)
Vocalization after facial action	0.42 (0.79)	0.30 (0.37)	1.08 (1.38)	1.02 (1.02)
Second face-to-face play				
Vocalization and smile	8.83 (8.89)	8.56 (8.52)	10.50 (12.32)	9.28 (8.87)
Vocalization and frown	7.00 (10.37)	2.61 (3.27)	8.25 (12.81)	4.99 (6.79)
Vocalization in facial action	9.33 (9.90)	5.76 (5.42)	12.75 (13.99)	8.09 (6.94)
Facial action in vocalization	1.25 (1.82)	0.77 (0.82)	0.42 (1.16)	0.71 (0.86)
Vocalization before facial action	2.92 (2.54)	2.28 (2.02)	3.00 (2.73)	2.75 (2.80)
Vocalization after facial action	2.33 (2.31)	2.36 (2.22)	2.58 (2.75)	2.74 (2.80)

Note. $N = 12$. Standard deviations are in parentheses. Vocalization and smile = vocalizations that were temporally associated with smile displays; vocalization and frown = vocalizations that were temporally associated with anger–distress–sadness displays.

facial action ended before the vocalization, $F(1, 11) = 11.48$, $p = .006$.[3] No other main effects or interactions were significant.

Thus, the sequence types indicating the degree of coordination most different from chance involved the following two scenarios: (a) When two communicative events were temporally associated with one another, one communicative event (either the vocalization or the coded facial action) was temporally embedded in the other, and (b) when a vocalization occurred with a coded facial action, the vocalization tended to end before the coded facial action, regardless of which communicative event started first.

Discussion

This study provides quantitative evidence that infants systematically sequence communicative actions involving vocalizations and facial expressions within the first 6 months of life. Our results confirm earlier reports of coordination based on time-based data in that sequences of coordination occurred more than expected by chance. Using an event-based analysis, we were able to examine what type of coordinated sequences occurred. Specifically, coordinated events produced by the infants were organized by two principles: (a) When two communicative events were temporally associated, one event (either the vocalization or the coded facial action) tended to be temporally embedded within the other, and (b) when a vocalization was temporally associated with a coded facial action, the vocalization tended to end before the coded facial action. In addition, coordination was most prevalent in anger–distress facial expressions but also occurred significantly with smiling facial expressions. The research revealed no systematic tendency for the observed patterns of coordination to vary with age or protocol episode, although the sample sizes may have been too small to yield a powerful test of developmental change.

The Importance of Coordinated Action and the Use of Event-Based Analyses

A number of communicative functions may be served when infants coordinate actions across two modalities. First, the coordination of events may provide a stronger or more compelling signal

[3] The vocalization-ending-first principle cannot be accounted for by the fact that vocalizations have shorter mean durations than facial actions. If the analysis were biased by short durations of vocalizations, one would expect the vocalization-before-facial-action and vocalization-after-facial-action sequence types to be equally likely and the likelihood of facial-action-in-vocalization sequences to be minimal. This, however, was not the case (see Figure 3).

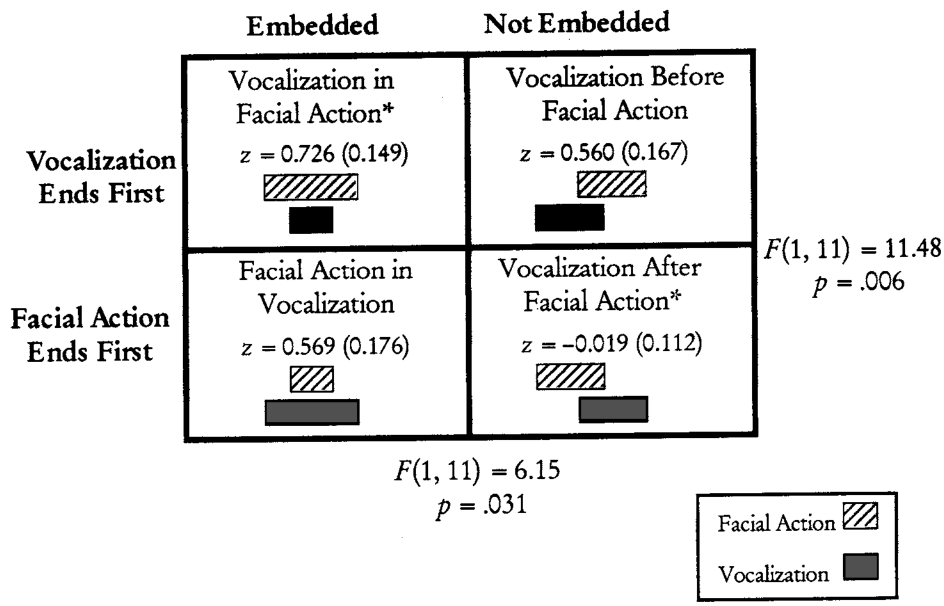

Figure 3. Follow-up analyses for the specific sequencing types of coordinated events. Standard deviations are in parentheses. Asterisks denote that the means are significantly different at the .05 level.

to the communicative partner than a message presented in either modality alone. That is, coordination may increase the quantity of signals. Second, one of the modalities involved in the coordinated event may reinforce or provide a level of clarity about the communicative or affective intent of the other modality. In other words, coordination may increase the specificity of the signal. Third, coordination in infants may lay a foundation for later complexity in child and adult communication.

Our results, like those of Fullmer and Messinger (1997) and Weinberg and Tronick (1994), indicate that infants do coordinate actions across the vocal and facial domains. However, the knowledge concerning coordination that is gained from time-based studies (that vocalizations and facial actions tend to overlap) is incomplete. The communicative message of overlapping events is likely to depend on their temporal ordering. An event-based method allowed us to examine such orderings.

Principles of Sequential Coordination

The communicative functions of coordinating across modalities may be served by embedding one within the other. When a message begins in one modality, the introduction of a communicative signal from a second modality may increase the clarity (specificity) or emphasis (quantity) of the message. The message then ends in the original modality, creating a coordinated signal that may be both qualitatively and quantitatively different than one sent in either one modality or the other. In these cases, the embedded event might be viewed as the focal point that accentuates the communicative signal of the temporally broader background event. The two sequence types that shared this embedded quality (vocalization in facial action and facial action in vocalization) produced the degree of coordination most different from chance occurrence.

At young ages, facial actions may serve as a particularly potent communicative modality, perhaps because facial actions provide clear indications of affective state. This may explain why infants tend to end vocalizations before facial actions when they produce coordinated events. Infants may tend to end their messages with facial actions rather than with vocalizations because vocalizations may tend to draw attention to the communicative signal, whereas the facial action leaves the recipient with a clearer affective message.

The optimal case of coordination in infant communication may be provided by a combination of the two principles—embedding one communicative event in the other and having the vocalization end before the facial action. This, in fact, is the description of the vocalization-in-facial-action sequence type, which resulted in the highest mean z score in the study. Facial actions may provide more affective information, providing a natural frame for communicative events, whereas vocalizations may be superior for getting another's attention (i.e., serving to clarify or emphasize). Circumstances embodying the opposite of the two principles—no embedding of one event in the other, with the facial action ending before the vocalization—are consistent with the vocalization-after-facial-action sequence type. This type had the lowest mean z score in the study, a score that was in fact negative, indicating that infants tended not to begin a communication with a facial action and end it with a vocalization.

The Bootstrapping Procedure

Prior research has not, to our knowledge, ever directly attempted to determine whether action events in infancy are associated nonrandomly across different modalities of action. The present research offers the first statistically validated indication that coordinated communication of specific temporal patterning does occur across facial and vocal modalities during infancy. The use of a bootstrapping analysis has also offered a solution to chronic problems in determining quantitative indexes of coordination. Previous analyses of coordination in time-based approaches were typically accomplished by using chi-square analyses to determine observed versus expected occurrences. These statistical procedures assume independence of units, in this case, fixed time intervals. However, the fixed time intervals examined in a time-based approach are not truly independent because actions within successive time intervals are likely to be correlated (but see Bakeman & Dorval, 1989). Bootstrapping circumvents this issue. Observed versus expected outcomes are compared without assuming the independence of time intervals when the degree of coordination expected by chance is calculated.

Limitations and Evidence for Future Investigations

Even with a relatively small sample size of 12 infants, there was evidence of significant coordination across communicative modalities analyzed from an event-based perspective. However, these results may only be generalizable to infants fitting the demographic profiles of those studied here. Replication with a larger sample size would enable examination of the vocalization–facial action events in greater detail and could provide support for the stability of these findings, especially across different socioeconomic and ethnic groups. The small sample size also prevented the analysis of gender as a variable in the present study. Coordination of vocalizations or facial actions may be different in boys and girls; however, a larger sample would be necessary to make proper conclusions.

Individual differences are apparent. The nonsignificant episode effect is probably due to large individual differences in the degree of coordination. Data from a larger participant population might reveal significant episode differences. It may also be that the tendency to coordinate actions across communicative modalities is robust despite the change in caregivers' behavior during the modified still-face episode. That is, the tendency to coordinate across modalities may not have been affected by an interruption in the normal flow of interaction in the modified still-face episode.

The sequential organization of facial actions and vocalizations did not differ by age. The data indicate that infants already show systematic coordination across communicative modalities by 3 months of age. Future investigations should examine sequentially coordinated communicative events both in younger infants (to find the youngest age at which infants coordinate action) and in older infants (to obtain information regarding longer term developmental trends). Finally, with event-based analyses, investigators can now look at the parent–infant dyad as the unit of analysis and examine how the parent and the infant sequentially coordinate their respective communicative signals.

The logical next step would be to analyze the extent to which coordination occurs between the specific vocalization types (e.g., complaints, noncomplaints) and the specific facial actions (e.g., smiles, grimaces, distress behaviors). It would be expected from time-based studies (Fullmer & Messinger, 1997; Weinberg & Tronick, 1994) that positive–neutral vocalizations would tend to be coordinated with positive facial actions and negative vocalizations with negative facial actions. Analysis from an event-based perspective would provide evidence concerning the degree to which coordination as well as the sequence patterns between these individual positive and negative communicative actions occur.

At present, the event-based analysis as described in the present study appears to be most effective in situations where coordination involves two modalities. These modalities do not have to involve intrasubject variables (i.e., infant vocalizations and infant facial actions) but can also involve intersubject variables (i.e., infant smiles and mother smiles). It should also be noted that as the number of coding categories in each modality increases, the resulting coordination patterns also increase in number and complexity. Furthermore, we also modified the bootstrapping procedure to incorporate more than two modalities into the event-based analysis.

Conclusion

This study provides clear evidence that coordination across communicative modalities can be examined from an event-based perspective, confirms certain results from other coordination studies that used time-based approaches, outlines new methods for analysis of event-based data, and uses these new methods to reveal novel features of infant–caregiver communication. Infants as young as 3 and 6 months of age systematically coordinated vocalizations and facial actions across interactive contexts and appeared to be following two principles of coordination. Infants were particularly likely to temporally embed one communicative event in another and, in addition, were particularly likely to end vocalizations before facial actions. The event-based approach avoided methodological issues introduced by traditional time-based analyses and satisfied ecological validity by conforming to the apparent inclination of real observers to respond to communicative events rather than arbitrary, fixed time intervals. This study has also opened up the possibility of further event-based investigations that may contribute to the study of coordinated events in development.

References

Action Analysis Coding and Training (Version 2.1) [Computer software]. (1996). Miami, FL: Intelligent Hearing Systems.

Austin, J. L. (1962). *How to do things with words.* London: Oxford University Press.

Bakeman, R., & Dorval, B. (1989). The distinction between sampling independence and empirical independence in sequential analysis. *Behavioral Assessment, 11,* 31–37.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic Press.

Camras, L. A., Oster, H., Campos, J. J., Miyake, K., & Bradshaw, D. (1992). Japanese and American infants' responses to arm restraint. *Developmental Psychology, 28,* 578–583.

Cobo-Lewis, A. B. (1998). *Using the parametric bootstrap for event-based analysis of coordination.* Manuscript in preparation, University of Miami.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 10,* 37–46.

Demos, V. (1982). Facial expressions of infants and toddlers. In T. Field & A. Fogel (Eds.), *Emotion and early interaction* (pp. 39–74). Hillsdale, NJ: Erlbaum.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cold (Ed.), *Nebraska Symposium on Motivation* (pp. 207–283). Lincoln: University of Nebraska Press.

Ekman, P., & Friesen, W. (1978). *The Facial Action Coding System.* Palo Alto, CA: Consulting Psychologists Press.

Fullmer, C. E., & Messinger, D. S. (1997, April). *Coordination of infant behaviors in response to parental still-face.* Poster session presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24,* 95–112.

Kaye, K., & Fogel, A. (1980). The temporal structure of face-to-face communication between mothers and infants. *Developmental Psychology, 16,* 454–464.

Messinger, D. S. (1994). *The development of smiling: A dynamic systems approach.* Unpublished doctoral dissertation, University of Utah.

Mooney, C. Z., & Duval, R. D. (1993). *A nonparametric approach to statistical inference.* Newbury Park, CA: Sage.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Vol. 1. Production* (pp. 93–112). New York: Academic Press.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development, 59,* 441–449.

Oller, D. K., Yale, M. E., & Delgado, R. E. (1997, April). *Development of coordination across modalities of communication: Coding and analysis tools.* Poster session presented at the biennial meeting of the Society for Research in Child Development, Washington, DC.

Oster, H., & Rosenstein, D. (in press). *Baby FACS: Analyzing facial movement in infants.* Palo Alto, CA: Consulting Psychologists Press.

Tomkins, S. (1963). *Affect, imagery and consciousness.* New York: Springer.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry, 17,* 1–13.

van Beek, Y., Hopkins, B., & Hoeksma, J. B. (1994). Development of communicative behaviors in preterm infants: The effects of birthweight status and gestational age. *Infant Behavior and Development, 17,* 107–117.

Weinberg, M. K., & Tronick, E. Z. (1994). Beyond the face: An empirical study of infant affective configurations of facial, vocal, gestural, and regulatory behaviors. *Child Development, 65,* 1503–1515.

Received September 17, 1997
Revision received June 17, 1998
Accepted June 29, 1998 ■