Head Movement Dynamics During Normal and Perturbed Parent-Infant Interaction

Zakia Hammal[1], Jeffrey F. Cohn[1,2], Daniel S. Messinger[3], Whitney I. Mattson[3], & Mohammad H. Mahoor[4]
[1]Robotics Institute, Carnegie Mellon University, Pittsburgh, USA
[2]Department of Psychology, University of Pittsburgh, USA
[3]Department of Psychology, University of Miami, Miami, USA
[4]Department of Electrical and Computing Engineering, University of Denver, USA

Abstract— We investigated the dynamics of head motion in parents and infants during an age-appropriate, well-validated emotion induction, the Face-to-Face/Still-Face procedure. Participants were 12 ethnically diverse 6-month-old infants and their mother or father. During infant gaze toward the parent, infant angular amplitude and velocity of pitch and yaw decreased from face-to-face (FF) to still-face (SF) episodes and remained lower in the following Reunion. During infant gaze away from the parent, angular velocity of pitch decreased from FF to SF and remained lower in the Reunion (RE). Windowed cross-correlation suggested strong bidirectional effects with frequent shifts in the direction of influence. The number of significant positive and negative peaks was higher during FF than RE. Gaze toward and away from the parent was modestly predicted by head orientation. Together, these findings suggest that head motion is strongly related to age-appropriate emotion challenge, are consistent with the hypothesis that perturbations of normal responsiveness carry-over even after the parent resumes normal responsiveness in the reunion, and that there are frequent changes in direction of influence in the postural domain.

Keywords—Still Face Paradigm; Infancy; Parent-infant interaction; Head movements; Gaze; Social interaction.

I. INTRODUCTION

In everyday language and experience, we encounter associations between head motion and emotional expression. We lower our head in sorrow and raise it in pride. Yet, the relation between head motion and emotion is relatively neglected in the psychology of emotion literature, with few exceptions [1, 2]. Descriptions of prototypic emotions emphasize non-rigid deformations of facial features (e.g., laterally stretched lip corners in fear) rather than the context of rigid head motion in which they may occur (e.g., pulling the head back and away from a perceived threat).

Existing evidences suggest that head motion, facial expression of emotion, and attention may be closely coordinated. When an infant visually tracks an upwardly moving target, their brows raise; when they track the very same target moving from below their line of sight, their brows are lowered and pulled together. These co-occurrences appear to result from coordinated motor patterns [3]. In adult displays of embarrassment, smiles intensify as the head pitches down and often to the side [1, 4].

Head motion serves to regulate social interaction as well. Head nods and turns serve turn-taking [5] and back-channeling [6] functions and vary with the expressiveness of interactive partners. When one partner's expression becomes attenuated, the velocity of head nods and turns increase in the other partner [7]. Head motion also reflects individual differences, such as gender. Head nods and turns are greater in women than in men [8]. Realistic head motion is believed to be fundamental to the success of virtual human agents [9, 10].

In automated facial expression analysis, head motion has mostly been considered a "nuisance" variable: something to control for when extracting features for action unit or expression detection. It is standard to remove the confounding effects of held tilt, pitch, and zoom prior to analysis [11]. Few investigators in this area have considered the relation between head motion as a feature for facial expression or emotion detection. A notable exception is Gunes and Pantic [12] who found meaningful correlations between head movement and perceived dimensions of emotion (e.g., valence and arousal) in intelligent virtual agents.

As an initial step toward understanding the contribution of head motion to emotion communication and experience, we studied head motion in face-to-face interactions between parents and infants. We used a well-validated paradigm for emotion induction. The *Still-Face paradigm* [13, 14] consists of three episodes: *Normal* interaction in which the parent plays with their infant as they might at home; a Still-Face (*SF*) episode in which they remain unresponsive; and a *Reunion* (*RE*) episode in which the parent again becomes responsive and attempts to initiate play. Shared positive affect and co-regulation of social interaction are key developmental tasks in the first half-year; and infants respond powerfully to perturbations of typical parent behavior. During the normal interaction, infants show positive affect and well-coordinated patterns of play with their parent. In response to the SF, they often attempt to elicit the parent to respond; unsuccessful, they become withdrawn and affectively negative. This "still-face" effect is likely to continue into the subsequent reunion (see, for instance, [14]).

Figure. 1. Face-to-face interaction

Qualitative descriptions of the SF effect have often referred to head movement, such as when the infant looks away following an attempt to elicit positive affect from the parent by smiling. Tronick [13], for example, described a SF episode in which the infant initially orients toward the mother and greets her. When the mother fails to respond, the infant "alternates brief glances toward her with glances away from her." Finally, "as these attempts fail, the infant eventually withdraws, orients his face and body away from his mother, and stays turned from her."

Little is known beyond qualitative observations about head motion in this paradigm. When head motion has been coded, findings have not been reported separately. Head motion has been pooled with information from other modalities. For instance, Beebe and colleagues [15], separately coded facial expression, gaze, head movement, and vocalization, but then combined them into "ordinalized behavior scales". Messinger and colleagues [16] investigated coordination of infant gaze and facial expression but not head movement. Thus, little is known about the unique contribution of head motion in the coordination of emotion exchanges between parents and infants.

We used the SF paradigm to investigate this issue. Specifically, we hypothesized that head motion would show a still-face effect analogous to what has been reported for facial expression: decreased head motion in the still face relative to the normal interaction that is maintained in the following Reunion episode. Infants were 6 months of age, which is a prime period for face-to-face play. Individual differences in how infants respond to FF and SF at this age are predictive of a range of developmental outcomes [17, 18], including attachment security and behavior problems. We hypothesized that head movement and gaze in the SF paradigm would reveal the challenges of this age-appropriate emotion-inducing context. Specifically, we hypothesized that during FF, head movement would be greater in mothers and infants as they attempted to coordinate their affective engagement. During SF, head motion would be attenuated as infants became increasingly withdrawn. And in the Reunion, carryover of negative affect and withdrawal would serve to keep head motion depressed.

II. METHODS

A. Participants and Observational Procedures

Participants were 12 ethnically diverse 6-month-old infants (M = 6.20, SD = 0.43) and their parents (11 mothers, 1 father). Two infants were African-American, 2 Asian-American, 4 Hispanic-American, and 4 European-American. Eight were boys. All parents gave their informed consent to the procedures.

Infants were positioned in an infant-seat facing their parent who was seated in front of them (see Figure. 1). Following previous research [19], the procedure consisted of three contiguous episodes: 3 minutes of normal interaction (*FF*) where the parent plays with the infant as he/she might do at home; 2 minutes in which the parent remained oriented toward the infant but was unresponsive (Still-Face: *SF*); and 3 minutes in which the parent resumed normal play (Reunion: *RE*). No specific verbal, non-verbal, or behavioral instructions were given to the parents in how to play or elicit infant attention. A 2 seconds tone signaled the parent to begin and end each episode. If the infant became significantly distressed during the SF, (defined as crying for more than 30 seconds), the episode was terminated early. Consequently, the SF episodes ranged from 37 to 118 seconds (M = 115.19, SD =12.47).

The interaction was recorded using Sony DXC190 compact cameras at 29.97 f/s, and Countryman E6 Directional Earset Microphone and B6 Omnidirectional lavalier microphone. Face orientation to the cameras was approximately 20° from frontal for the infants and close to frontal for the parents, with considerable head motion (up to about 30° deviation from frontal faces) for each.

B. Manual Gaze Coding

Infant gaze toward and away from the parent was continuously coded offline by a trained observer. To assess reliability, a second trained observer randomly coded 25% percent of the videos. Chance-corrected agreement was quantified using coefficient kappa [20]. Individual video frames were the unit of analysis. The obtained kappa was 0.90, which represent high agreement. Parent gaze was not coded as parents were almost always looking toward their infant.

C. Automatic Head Tracking

The CISRO cylinder-based 3D head tracker was used to model the 6 degrees of freedom of rigid head motion [21]. This tracker has high concurrent validity with alternative methods and is capable of revealing dynamics of head motion in face-to-face interaction. For pitch and yaw in adults, the tracker correlated 0.93 and 0.94, respectively, with that of Xiao and colleagues [22]. The corresponding correlations with structure from motion estimates [23] were 0.96, and 0.93. In a recent application, the tracker significantly differentiated between low- and high conflict episodes in adult intimate and distressed couples [26].

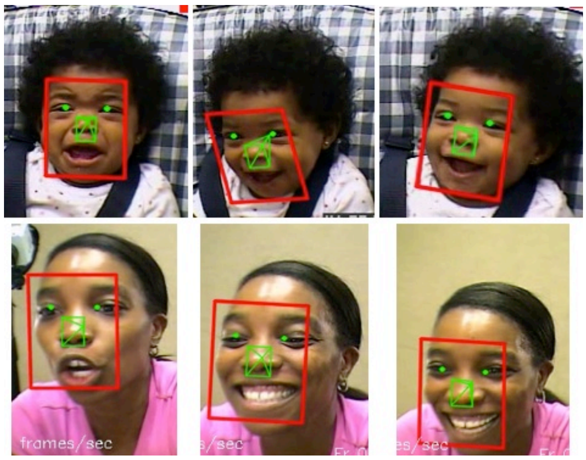

Figure. 2. Examples of tracking results of infant and mother during different moments of FF episode. Images within columns correspond to same moment for mother and infant.

For each participant (parent and infant), the tracker was initialized on a manually selected near-frontal image prior to tracking. Figure 2 shows examples of tracking results for a mother and her infant during a FF episode.

Two videos from one condition each in two dyads could not be tracked due to video corruption. For this reason, the two videos were omitted from the analyses reported below.

For each video frame, the tracker output 6 degrees of freedom of rigid head motion or a failure message when a frame could not be tracked. For parents, 6.75% of frames could not be tracked. The corresponding figure for infants was 41.08%. The tracker outputted eye position as well but without sufficient precision to infer direction of gaze.

To evaluate the quality of the tracking, we visually reviewed the tracking results overlaid on the video (see red boxes and green pyramids' orientation in Figure. 2). In 4% of the parents' video, visual review indicated errors in tracking. For infants, the corresponding figure was 9.72%. These frames were excluded from further analysis. Table 1 reports the distribution of tracked frames that met visual review for good tracking. Better performance was found for parents compared with infants, especially in the SF when head motion and occlusion were nearly absent for parents. For infants, better performance was found for FF and RE than for SF. Reasons for failure in infants included chewing on a shoulder strap, hand in mouth, extreme rapid head turn, and lack of clear boundaries between their lower jaw and chest. Training the tracker to achieve better performances for head movement of infants, which differ in multiple respects from those of adults, is a topic for future research.

D. *Measurement of Head Movement*

Angles of the head in the anterior–posterior and lateral directions were selected to assess head movement. These directions correspond to the meaningful motion of head nods (i.e. pitch) and head turns (i.e. yaw), respectively. Head angles (i.e. pitch and yaw) on a frame-by-frame basis were first converted into displacement and velocity. The displacement corresponds to the amplitude of movement for each selected head angle over the interaction segment. The velocity corresponds to the speed of the displacement for each head angle over the interaction segment from head angle sample for each frame. Velocity was computed as the first derivative of displacement.

TABLE 1. PROPORTION OF GOOD TRACKING FOR PARENTS AND INFANTS

Good Tracking	Episode		
	Face-to-Face	***Still Face***	***Reunion***
Mothers	82.5^{a}	99.3^{b}	85.7^{a}
Infants	57.8^{a}	35.3^{b}	54.4^{a}
Both mother and infant	43.5^{a}	34.5^{a}	41.3^{a}

Note. Within rows, means with different superscripts differ at $p < 0.05$. Within columns, scores for mothers were significantly higher than those for infants at $p < 0.05$.

Head displacement and velocity were converted into angular displacement and angular velocity by subtracting the mean overall head angle across the segment from head angle sample for each frame. We used the overall mean head angle, which afforded an estimate of the overall head position for each participant in each episode saving both the estimated displacement and velocity for each frame.

The root mean square (RMS) of the horizontal (RMS–yaw) and vertical (RMS–pitch) angular displacement and angular velocity were then calculated for each episode for each parent and infant.

So that missing data would not bias measurements, the RMSs of the angular displacement and angular velocity were computed for each valid segment, normalized by the duration of the segment and denoted *nRMS.* Then for each video the mean of the consecutive normalized RMSs over the sequence was computed. For the example shown in Figure. 3, we would first compute the normalized RMS for each consecutive segment: *nRMS1, nRMS2* and *nRMS3.* The RMS of the angular displacement of the entire videos would then be computed as the mean of these three normalized mRMSs = ((*nRMS1* + *nRMS2*+ *nRMS3*)/3).

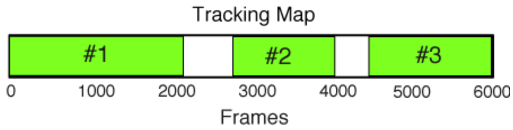

Figure. 3. Green segments correspond to good tracking; white to missing data or tracking failure.

III. RESULTS

Because of the repeated-measures nature of the data, Student's paired t-tests were used to test differences between episodes. The RMSs of the angular displacement and angular velocity for pitch and yaw were the primary dependent measures.

TABLE 2. MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) OF MANUALLY CODED INFANT GAZE TOWARD THE PARENT.

	Episode		
	Face-to-Face	***Still Face***	***Reunion***
Gaze toward parent	0.45^{a} (0.23)	0.17^{b} (0.16)	0.42^{a} (0.24)

Note. Means with different superscripts differ at $p < 0.05$.

A. *Infant Gaze Toward the Parent*

Means and standard deviations (in parentheses) of manually coded infant gaze toward the parent are reported in Table 2. One can see that infants looked toward their parent nearly half the time in the Face-to-Face and Reunion episodes and significantly less in the Still Face. In the latter, gaze toward the parent averaged only 17% of the time.

B. *Parent Head Movement*

As a validity check, we compared parent head movement across episodes. Because parents were instructed to remain unresponsive during the SF, we anticipated that their head movement would be less during the SF than during FF or RE. Parent head angular amplitude and angular velocity were consistent with this expectation as shown in Table 3 and Table 4.

C. *Infant Head Movement*

Because infant head motion may vary with gaze toward or away from the parent, head motion was analyzed separately for gaze toward and away from the parent.

1) *Infant gaze toward the parent*

During infant gaze toward the parent, both angular amplitude and angular velocity of pitch and yaw decreased significantly in SF compared with FF (see columns FF-SF in Table 3 and Table 4). The obtained results are consistent with previous findings that gazing at the parent, smiling, and social bidding (smiling while gazing at the parent) decreased with time in the SF relative to FF. Head angular amplitude and angular velocity were also lower in RE relative to FF (see columns FF-RE in Table 3 and Table 4). However, SF and RE did not significantly differ. Indeed, even if both angular amplitude and angular velocity of pitch and yaw increased from SF to RE (see columns SF-RE in Table 3 and Table 4) they did not change significantly (p>0.05), which suggests a carryover effect from SF to RE.

2) *Infant gazes away from the parent*

During infant gaze away from the parent, no differences were found for angular amplitude. Angular velocity decreased significantly from FF to SF and consistent with carry-over effect remained low during RE (Table 5).

D. *Windowed Cross Correlation and Peack Picking*

1) *Windowed Cross-Correlation*

To investigate whether parent-infant synchrony varied between Face-to-Face and Reunion, we used windowed cross-correlation and peak picking [25]. The windowed cross correlation produces positive and negative correlation values between two signals over time for each (time, lag) pair of values. The windowed cross correlation uses a temporally defined window to calculate successive local zero-order correlations over the course of an interaction.

TABLE 3. PITCH AND YAW ANGULAR AMPLITUDES FOR PARENTS AND INFANTS DURING INFANT GAZE TOWARD THE PARENT

		Means			Differences between episodes by paired *t*-test		
Amplitude		**FF**	**SF**	**RE**	**FF - SF**	**FF - RE**	**SF - RE**
Pitch	Parents	0.0014	3.6310e-04	0.0014	t = 4.04**	t = 0.06	t = -2.80**
	Infants	0.0023	9.4560e-04	0.0013	t = 2.24**	t = 2.09*	t = -0.89
Yaw	Parents	0.0014	3.6641e-04	0.0014	t = 4.85**	t = 0.04	t = -3.62**
	Infants	0.0023	0.0010	0.0013	t= 2.23**	t = 1.91*	t = -0.75

*Note. ** p < =0.05, * p < 0.10*

TABLE 4. PITCH AND YAW ANGULAR VELOCITIES FOR PARENTS AND INFANTS DURING INFANT GAZE TOWARD THE PARENT

		Means			Differences between episodes by paired *t*-test		
Velocity		**FF**	**SF**	**RE**	**FF - SF**	**FF - RE**	**SF - RE**
Pitch	Parents	0.0016	3.7175e-04	0.0016	t = 2.33**	t = -0.02	t = -1.89*
	Infants	0.0033	9.0071e-04	0.0014	t = 2.44**	t = 2.29**	t = -0.81
Yaw	Parents	0.0015	3.7147e-04	0.0014	t = 2.60**	t = 0.03	t = -2.22**
	Infants	0.0033	0.0011	0.0015	t = 2.32**	t = 1.95*	t = -0.71

*Note. ** p < 0.05, * p < 0.10*

TABLE 5. PITCH AND YAW ANGULAR AMPLITUDES FOR PARENTS AND INFANTS DURING INFANT GAZE AWAY FROM THE PARENT

		Means			Differences between episodes by paired *t*-test		
Amplitude		**FF**	**SF**	**RE**	**FF - SF**	**FF - RE**	**SF - RE**
Pitch	Parents	0.0017	0.0003	0.0019	t = 3.06*	t = -0.53	t = -4.11*
	Infants	0.0027	0.0019	0.0021	t = 1.52	t = 1.09	t = -0.18
Yaw	Parents	0.0017	0.0003	0.0020	t = 3.30*	t = -0.59	t = -3.87 *
	Infants	0.0027	0.0018	0.0020	t =1.76	t = 1.42	t = -0.26

*Note. * p <= 0.01*

TABLE 6. PITCH AND YAW ANGULAR VELOCITIES FOR PARENTS AND INFANTS DURING INFANT GAZE AWAY FROM THE PARENT

		Means			Differences between episodes by paired *t*-test		
Velocity		**FF**	**SF**	**RE**	**FF - SF**	**FF - RE**	**SF - RE**
Pitch	Parents	0.0021	0.0003	0.0025	t = 2.08*	t = -0.79	t = -2.98 ***
	Infants	0.0040	0.0023	0.0030	t = 1.99*	t = 0.96	t = -0.55
Yaw	Parents	0.0021	0.0003	0.0026	t = 2.18**	t = -0.77	t = -2.75**
	Infants	0.0038	0.0021	0.0028	t = 2.17**	t = 1.17	t = -0.62

*Note. *** p<=0.01, ** p <= 0.05, * p < 0.10*

The windowed cross correlation involves the selection of a set of parameters: the sliding windows size ($w_{\max}$), the window increment (w_{inc}), maximum lag ($\tau_{\max}$) and the lag increment (τ_{inc}). Plus or minus 2s appeared to be meaningful for perception of head nods (i.e. pitch change) and turns (i.e. yaw). We therefore set window size $w_{\max}$ = 4 seconds and increment window w_{inc}= 1. The maximum lag $\tau_{\max}$ is the maximum interval of time that separates the beginning of the two windows. The maximum lag should allow capture of the possible lags of interest and exclusion of lags that are not of interest. After exploring different lag values, we choose a maximum lag $\tau_{\max}$=1/2 seconds x 29.97 sample/sec = 15 samples. Using these parameter values, the windowed cross-correlation for pitch and yaw was computed for the three episodes: FF, SF and RE.

For each dyad, the windowed cross correlation measures the correlations between the previous head movements of the parents and the subsequent head movements of the infants. In the current study, the windowed cross correlation was used to measure the correlations between the amplitude as well as the velocity of the parents' head movements and the amplitude and the velocity of the head movements of their infants. Figure. 4 shows example of windowed cross correlation for pitch amplitude and velocity for one dyad during the FF episode.

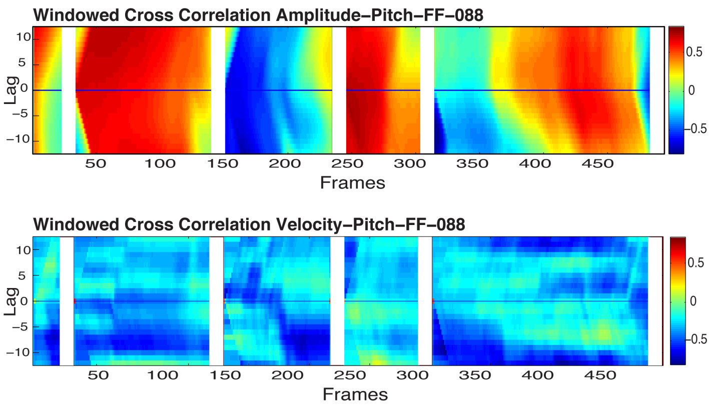

Figure. 4. Windowed cross correlation for pitch amplitude (top) and velocity (bottom).

The obtained rows highlight local periods of positive (red plots) and negative correlations (blue plots) between the amplitude and velocity of head movements of the parent and his infant (see Figure. 4). One can observe that the direction of the correlations changes dynamically over time. Head amplitude of the parent predicts head amplitude of the infant (characterized by positive or negative correlation in the area above the plot midline), and infant predicts head amplitude of the parent (characterized by positive or negative correlation in the area below the plot midline). Overall, at any given time parents and infants form a coregulated dyadic head pose system. Each partner's head movements predict the other in time and infant to parent influence is mirrored by parent to infant influence.

2) Peak Picking

To better analyze the patterns of change in the obtained cross correlation matrix, [25] proposed a peak selection algorithm that selects the peak correlation at each elapsed time according to flexible criteria. The second line of Figure. 5 shows an example of the peaks of correlation (red represents positive peaks and blue negative peaks) obtained using the peak-picking algorithm on the correlation matrix reported in the first line of Figure. 5. The obtained results show that the direction of the correlations changes dynamically over time. The pattern of association between parents and infants is non-stationary with frequent changes in which partner is leading the other.

So that low correlations would not bias the findings, only correlations greater than 0.5 were included for peak picking results analyses. Correlations at this threshold or higher represent effect sizes of moderate or larger.

Student's paired t-tests were used to test differences in head movements' correlations of parents and infants between episodes. To do so, similarly to the RMSs (see section II.D), so that missing data would not bias measurements, the number of positive and negative peaks for each episode was computed for each valid segment separately and normalized by the duration of that segment. The means of the normalized peaks are then computed and used as the primary dependent measures. As parents were asked to stop interaction with their infants in SF episode, only paired comparisons of positive and negative peaks between FF and RE episodes were conducted.

The number of correlation peaks decreased from FF to RE. The number of positive and negative peaks decreased from FF to RE for pitch amplitude (t= 2.58, p=0.02 and t= 2, p=0.07, respectively) and yaw amplitude (t= 2.36, p=0.03 and t= 1.81, p=0.09, respectively). The number of correlation peaks decreased as well from FF to RE for pitch velocity for positive (t= 11.7, p=3.68e-07) and negative peaks (t= 17.55, p= 6.75e-11). The number of correlation peaks decreased as well from FF to RE for yaw velocity for positive peaks (t=2.21, p=0.05). These results are consistent with the hypothesis that the SF experience carries over into the RE episode. The infants are still affected by the parents' lack of responsiveness and parents and infants do not coordinate at pre-perturbation levels.

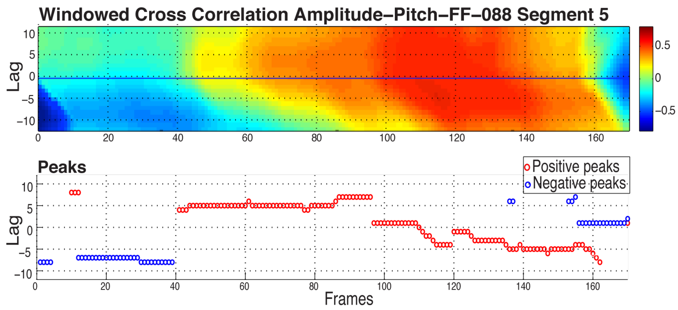

Figure. 5. For the windowed cross-correlation plot in the upper panel, we computed correlation peaks using the peak-picking algorithm. The peaks are shown in the lower panel. Red indicates positive correlation; blue indicates negative correlation. When peak correlations are positive, mother and infant increase or decrease their head movement in phase (i.e. synchronously). When peak correlations are negative, they do same movement but out of phase (i.e., mother or infant follow the other).

E. Correlation Between Head orientation and Gaze

As an additional analysis, we investigated the extent to which gaze could be estimated from head orientation. Both head orientation and gaze are coordinated by the same brain circuits, the medial longitudinal fasciculus in the brainstem proximal to the facial nerve nucleus [26]. We therefore anticipated that head orientation could be useful in estimating gaze.

Pitch, yaw, and gaze were moderately correlated in all three episodes (FF, SF, and RE, see Figure. 6). To evaluate the extent to which gaze could be predicted from pitch and yaw, we first normalized pitch and yaw with respect to a frontal pose (i.e. orientation toward the parent). The absolute value of pitch and yaw then entered into a logistic regression to predict gaze (toward or away from the parent).

Pitch and yaw proved modest predictors of gaze. The Hosmer-Lemshow test was highly significant, $\chi^2 = 1915.88$, $df = 8$, $p < 0.001$. Gaze was correctly predicted in 65.2% of frames, kappa = 0.273.

IV. DISCUSSION

Parent and infant head motion showed strong responses to the emotion challenges presented by the SF paradigm. During the FF episode, in which parent and infant often coordinate bouts of positive interaction, infant head amplitude was greater and faster in comparison with the SF. Infants responded to the SF with decreased amplitude and velocity. Strikingly, this effect carried over, or continued, into the ensuing RE. Even when the parent resumed normal responsiveness and affect, the infant remained more withdrawn, which is supported by the greatly reduced frequency of significant peaks in the SF windowed correlations and reduced number of changes in phase. The inertia we observed from the SF into the RE was of course short lived. However, were such unresponsiveness typical of the infants' experience, we would expect that the infant's adaptation would become increasingly resistant to change.

Previous work in emotion emphasizes facial expression. The face is, in fact, a rich source of information about emotion and intention. Head movement may not rival facial expression in subtlety of expression, but it conveys much about the experience of parent and infant. When infant was gazing at parent their head motion changed markedly when parent became unresponsive. That effect not only was dramatic but had consequence for subsequent Reunion. The temporal coordination of head movement between parent and infant strongly suggests a deep inter-subjectivity between them. Body language such as this is well worth our attention for automatic affect recognition.

Parent and infant head motion were highly correlated with frequent shifts in the direction of influence. One interpretation is that this pattern represents a process of mismatch and repair described by Tronick and colleagues [27]. They found that joint states of positive engagement became disrupted, presenting a challenge for the dyad to resolve. The frequent changes in phase we observed may reflect this process. Alternatively, frequent shifts in phase may reflect playfulness or a kind of turn taking. Further work is needed that integrates temporal coordination of facial and also vocal expression with head motion.

Many approaches to bivariate time series analysis assume stationarity in the signals. Stationarity is normative in physical signals but not in social signals. We used windowed cross-correlation and peak picking to investigate the temporal coordination of parent and infant. These methods are, however, descriptive. There is a need for methods that can model the underlying dynamics of non-stationary social time series [28].

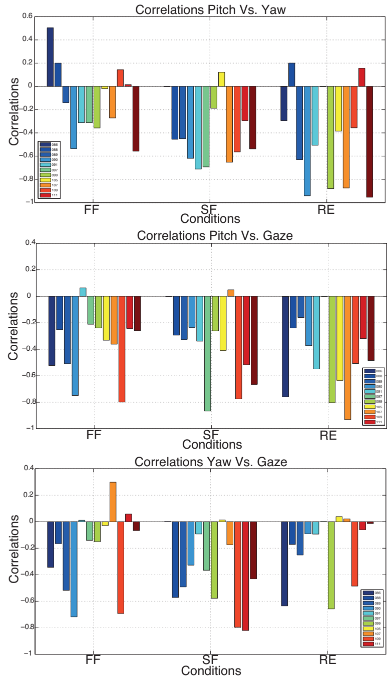

Figure. 6. Correlation between pitch and yaw, pitch and gaze, and yaw and gaze.

The current work opens new directions for future investigations of face-to-face interaction between parents and infants. One direction would be coordination between head motion and emotion exchanges between parents and infants. We currently are exploring the temporal coordination of head motion and facial expression in face-to-face interaction. Another direction is to explicitly include head motion as a feature to inform action unit and expression detection. Head motion is not simply a source of error (as in registration error), but a potential contributor to improved action unit detection. We aim to test this hypothesis. Yet another contribution is to inform human-robot and human-agent interaction. Head movement figures in behavioral mirroring and mimicry among people and qualifies the communicative function of facial actions (e.g., head pitch in embarrassment). Social robots that can interact naturally with humans will need the ability to use appropriately this important channel of communication. Social robots as well as virtual human agents would benefit from further research into human-human interaction with respect to head motion and other body language. Work in this area is just beginning.

V. CONCLUSION

In summary, we found that head movement is a powerful signal in social interaction that strongly varies with context. When the parent was responsive, high amplitude, high velocity, and temporal coordination of parent-infant head movement was achieved. Parent unresponsiveness in contrast undermined this behavioral organization. The infant became withdrawn and unresponsive. This attenuated responsiveness carried over even after the parent resumed normal behavior.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grants BCS 1052603, BCS 1052736, and 1052781 and by NIH grant HD047417. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Institute of Health or the National Science Foundation.

REFERENCES

[1] D. Keltner, "Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement and shame," *Journal of Personality and Social Psychology,* vol. 68, pp. 441-454, 1995.

[2] J. F. Cohn, L. I. Reed, T. Moriyama, J. Xiao, K. L. Schmidt, and Z. Ambadar, "Multimodal coordination of facial action, head rotation, and eye motion," presented at the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 2004.

[3] G. F. Michel, L. Camras, and J. Sullivan, "Infant interest expressions as coordinative motor structures," *Infant Behavior & Development,* vol. 15, pp. 347-358, 1992.

[4] Z. Ambadar, J. F. Cohn, and L. I. Reed, "All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous," *Journal of Nonverbal Behavior,* vol. 33, pp. 17-34, 2009.

[5] S. Duncan, "Some signals and rules for taking speaking turns in conversations," *Journal of Personality & Social Psychology,* vol. 23, pp. 283-292, 1972.

[6] M. L. Knapp and J. A. Hall, *Nonverbal behavior in human communication*, 7th ed. Boston: Wadsworth/Cengage, 2010.

[7] S. M. Boker, J. F. Cohn, B. J. Theobald, I. Matthews, J. Spies, and T. Brick, "Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars," *Philosophical Transactions B of the Royal Society,* vol. 364, pp. 3485–3495, 2009.

[8] S. M. Boker, J. F. Cohn, B. J. Theobald, I. Matthews, M. Mangini, J. R. Spies*, et al.*, "Something in the way we move: Motion, not perceived sex, influences nods in conversation," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 37, pp. 874-891, 2011.

[9] J. Gratch, N. Wang, A. Okhmatovskaia, F. Lamothe, Mathieu Morales, R. J. van der Werf*, et al.*, "Can virtual humans be more engaging than real ones?," presented at the International Conference on Human-Computer Interaction, Beijing, China, 2007.

[10] R. Li, T. S. Huang, and M. Danielsen, "Real-time multimodal human-avatar interaction," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 18, pp. 467-477, 2008.

[11] F. De la Torre and J. F. Cohn, "Visual analysis of humans: Facial expression analysis," in *Visual analysis of humans: Looking at people*, T. B. Moeslund, A. Hilton, A. U. Volker Krüger, and L. Sigal, Eds., ed: Springer, 2011, pp. 377-410.

[12] H. Gunes and M. Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *Intelligent virtual agents*. vol. 6356, J. A. e. al., Ed., ed Heidelberg: Springer-Verlag Berlin, 2010, pp. 371-377.

[13] E. Tronick, H. Als, L. Adamson, B. A. Susan Wise, and T. B. Brazelton, "The infant's response to entrapment between contradictory messages in face-to-face interaction," *Journal of the American Acadamy of Child Adolescent Psychiatry,* vol. 17, pp. 1-13, 1978.

[14] L. B. Adamson and J. E. Frick, "The still face: A history of a shared experimental paradigm," *Infancy,* vol. 4, pp. 451-473, 2003.

[15] B. Beebe, J. Jaffe, S. Markese, K. Buck, H. Chen, P. Cohen*, et al.*, "The origins of 12-month attachment: A microanalysis of 4-month mother-infant interaction," *Attachment & Human Development,* vol. 12, pp. 3-141 2010.

[16] D. S. Messinger, W. I. Mattson, M. H. Mohammad, and J. F. Cohn, "The eyes have it: Making positive expressions more positive and negative expressions more negative," *Emotion,* vol. 12, pp. 430-436, 2012.

[17] J. Jaffe, B. Beebe, S. Feldstein, C. L. Crown, and M. Jasnow, "Rhythms of dialogue in early infancy," *Monographs of the Society for Research in Child Development,* vol. 66, 2001.

[18] G. A. Moore, J. F. Cohn, and S. B. Campbell, "Infant affective responses to mother's still-face at 6 months differentially predict externalizing and internalizing behaviors at 18 months. ," *Developmental Psychology,* vol. 37, pp. 706-714, 2001.

[19] L. B. Adamson and J. E. Frick, "The still-face: A history of a shared experimental paradigm," *Infancy,* vol. 4, pp. 451-474, 2003.

[20] J. L. Fleiss, *Statistical methods for rates and proportions*. New York: Wiley, 1981.

[21] M. Cox, J. Nuevo-Chiquero, J. M. Saragih, and S. Lucey, "CSIRO Face Analysis SDK," ed. Brisbane, Australia, 2013.

[22] J.-S. Jang and T. Kanade, "Robust 3D head tracking by online feature registration," in *IEEE International Conference on Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, 2008.

[23] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-time combined 2D+3D active appearance models," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* pp. 535-542, 2004.

[25] S. M. Boker, J. L. Rotondo, M. Xu, and K. King, "Windowed cross–correlation and peak picking for the analysis of variability in the association between behavioral time series," *Psychological Methods,* vol. 7, pp. 338-355, 2002.

[26] E. M. Klier, W. Hongying, and J. D. Crawford, "Three-dimensional eye-head coordination is implemented downstream from the superior colliculus," *Journal of Neuro-physiology,* vol. 89, pp. 2839-2853, 2003.

[27] E. Z. Tronick and J. F. Cohn, "Infant-mother face-to-face interaction: Age and gender differences in coordination and the occurrence of miscoordination," *Child Development,* vol. 60, pp. 85-92, 1989.

[28] S. Chow, J.D. Haltigan and D.S. Messinger, D.S. (2010). "Dynamic Infant-Parent Affect Coupling during the Face-to-Face/Still-Face". Emotion, vol. 10, pp. 101-114.