Kinesic Credibility Assessment during Criminal Interviews

CITeR Final Report
October 2009

Matthew Jensen, Judee Burgoon, Stanley Slowik, Pete Blair and Dimitris Metaxas

University of Oklahoma, University of Arizona, Stanley Slowik, Inc., Texas State University, and Rutgers University

Problem

- **Kinesic analysis has shown promise in improving unobtrusive credibility assessment**
  - Number, duration and expansiveness of semantically meaningful gesture (illustrators)
  - Number, duration, and intensity of self-adaptors
  - Head movement and blink behavior
  - Asymmetric posture
- **Will kinesic analysis be able to discriminate between truth and deception under high-stakes?**

Proposed Milestones and Deliverables

Milestone	Description and Deliverable	Timeframe
(1) Acquire IRB approval	**Work with the IRB to ensure protection of suspects whose behavior is being examined**	**Completed**
(2) Prepare and segment lines of questioning	**Manually segment each interview according to lines of questions (primary question, follow-up questions); Extract video segment for kinesic processing**	**Completed**
(3) Process the segments with kinesic analysis	**Automatically process and extract kinesic features from each segment of the interviews; Manually review results to ensure accuracy**	**Completed**
(4) Test multiple classification techniques	**Use multiple techniques to identify most diagnostic combination of kinesic cues; Create classification models for high-stakes deception**	**Completed**

Description of the data

- **Suspects are interviewed as part of a crime investigation**
  - Assault
  - Rape
  - Manslaughter
- **Suspects are interviewed using a kinesic interviewing protocol developed by Stan Slowik**
  - Based on the BAI developed by Reid & Associates
- **Some suspects are innocent of the crime; others are guilty**
- **Ground truth is given by court outcome, evidence, judgment of interviewer**

Description of the data

- **Interviews cover narrative of crime and questions related to interviewee affective state**
- **Questions analyzed in this dataset:**

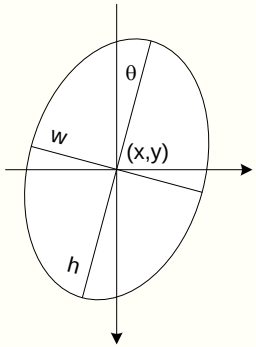
1. What kind of person is the victim?
2. Why might someone want to do this to the victim?
3. How do you feel about the accusation?
4. What do you think should happen to the person who did this?
5. How do you think the person who did this feels?

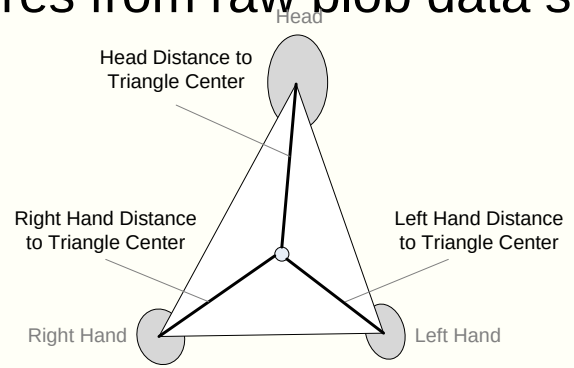
Operational Data Issues

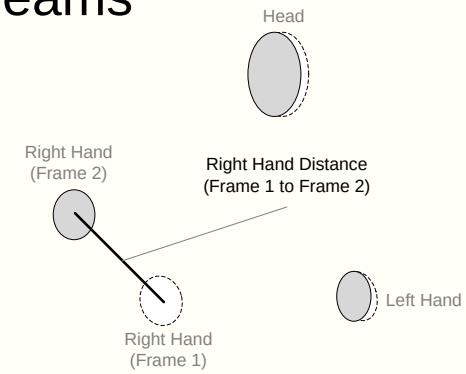
- **Poor lighting**
  - Shadows obscure facial and gesture features
  - Mixture of interior and exterior lighting
- **Occlusion**
  - Desks, chairs and other furniture
  - Beards, long hair, jackets
  - Hands in pockets
- **Suspect position**
  - Orientation away from the camera
  - Variable orientation (e.g., a swivel chair)

Blob Tracking

- **Track head and hands throughout a video segment**
  - Derive features from raw blob data streams







  - Feature values summarized through means and standard deviations
  - Significant occlusion of hands and quality of the video allowed only 18 videos to be analyzed (11 guilty and 7 truthful)
  - Limits the number of features we can analyze

Blobs - Findings

- **None of the adaptor and illustrator gesture features were significant in between-subjects tests**
  - Repeated measures analysis included distance, triangle and displacement features for both hands and the head
  - Some question effects, but nothing consistent across LH or RH

Transformed Variable: Average

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	TriCenter_RH	1377994.866	1	1377994.866	116.664	.000	.879	116.664	1.000
	TriCenter_LH	1458964.395	1	1458964.395	222.343	.000	.933	222.343	1.000
	Head_RH	4374873.048	1	4374873.048	204.430	.000	.927	204.430	1.000
	Head_LH	4647644.803	1	4647644.803	505.090	.000	.969	505.090	1.000
	RH_Diff	311.734	1	311.734	16.410	.001	.506	16.410	.967
	LH_Diff	395.141	1	395.141	16.427	.001	.507	16.427	.967
Guilty	TriCenter_RH	14135.467	1	14135.467	1.197	.290	.070	1.197	.177
	TriCenter_LH	12047.632	1	12047.632	1.836	.194	.103	1.836	.247
	Head_RH	947.269	1	947.269	.044	.836	.003	.044	.055
	Head_LH	1800.888	1	1800.888	.196	.664	.012	.196	.070
	RH_Diff	2.266	1	2.266	.119	.734	.007	.119	.062
	LH_Diff	10.537	1	10.537	.438	.517	.027	.438	.096
Error	TriCenter_RH	188986.098	16	11811.631					
	TriCenter_LH	104988.175	16	6561.761					
	Head_RH	342405.248	16	21400.328					
	Head_LH	147225.800	16	9201.613					
	RH_Diff	303.937	16	18.996					
	LH_Diff	384.869	16	24.054					

Blobs - Findings

- **Logistic Regression**
  - DV Guilt
  - IVs: 8 features capturing adaptor and illustrator gesturing
  - No individual IV is significantly diagnostic
  - Together they provide some diagnostic power

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	24.057	8	.002
	Block	24.057	8	.002
	Model	24.057	8	.002

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.000[a]	.737	1.000

Blobs - Findings

- **Logistic Regression Classification Accuracy**
  - ZeroR 61.1% (Classify all 18 cases as deceptive)
  - Logistic regression model produces 100% accuracy

Classification Table[a]

Observed			Predicted		
			Guilty		
			Guilty	Innocent	Percentage Correct
Step 1	Guilty	Guilty	11	0	100.0
		Innocent	0	7	100.0
		Overall Percentage			100.0

a. The cut value is .500

- **Questions about generalizability**
  - During 10-fold cross validation accuracy falls to 61.1%
  - Deceptive 54.5%; Truthful 71.4%

Active Shape Model facial landmark tracking

- **Tracks points of the face throughout a segment**
  - Identify blinks, head nods, head shakes
  - Counts and duration
- **All features are normalized according to the length of the segment**
- **Features from 31 videos were extracted**
  - 17 guilty
  - 14 innocent

Blinks, Nods, and Shakes - Findings

- **Repeated measures analysis with counts and durations for blinks, nods, and shakes with question as the repeated factor**
- **Nods, shakes and question factors not significant**
- **Guilty suspects exhibited a higher frequency of blinks ($F(1, 16) = 5.42$; $p = .033$)**
- **Not all of the questions were asked of all suspects**
  - This causes listwise deletion of cases in RM analysis
  - N = 18

Blinks, Nods, and Shakes - Findings

- **Missing values replaced with means to allow RM analysis (N = 31)**
- **Question factor, blink duration are significant (p<.1)**

Univariate Tests

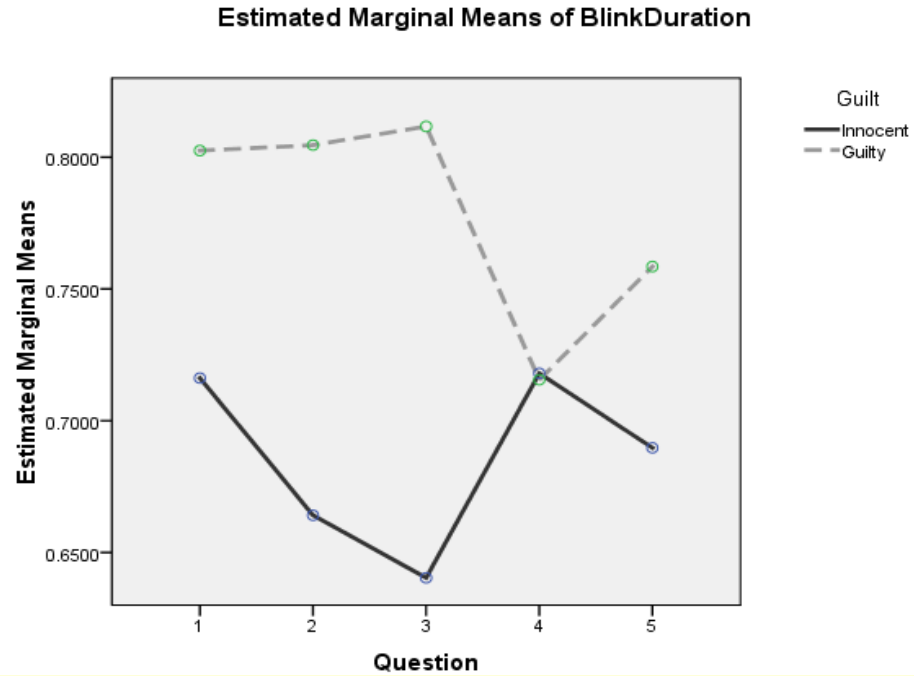
Measure		Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
BlinkCount	Contrast	6.038E-5	1	6.038E-5	1.303	.263	.043	1.303	.197
	Error	.001	29	4.634E-5					
BlinkDur	Contrast	.066	1	.066	3.398	.076	.105	3.398	.430
	Error	.565	29	.019					
NodCount	Contrast	1.784E-7	1	1.784E-7	.033	.857	.001	.033	.054
	Error	.000	29	5.400E-6					
NodDur	Contrast	5.242E-5	1	5.242E-5	.070	.794	.002	.070	.058
	Error	.022	29	.001					
ShakeCount	Contrast	2.831E-6	1	2.831E-6	.468	.499	.016	.468	.101
	Error	.000	29	6.049E-6					
ShakeDur	Contrast	7.263E-5	1	7.263E-5	.059	.810	.002	.059	.056
	Error	.036	29	.001					

The F tests the effect of Guilt. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

Blinks, Nods, and Shakes - Findings

- **Question effect on Blink Duration**
  - For blinks, “What should happen” question not diagnostic
  - Other items seem to produce diagnostic blink behavior



Blinks, Nods, and Shakes - Findings

- **Logistic Regression**
  - DV: Guilt
  - IVs: Blink durations from first 3 questions
  - No individual IV is significantly diagnostic
  - The overall model is not significant
  - Together they provide some diagnostic power

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	38.021[a]	.140	.187

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	4.664	3	.198
	Block	4.664	3	.198
	Model	4.664	3	.198

Blinks, Nods, and Shakes - Findings

- **Logistic Regression Classification Accuracy**
  - ZeroR 54.8% (Classify all 31 cases as deceptive)
  - Logistic regression model produces 74.2% accuracy

Classification Table[a]

Observed			Predicted Judgment D	Predicted Judgment T	Percentage Correct
Step 1	Judgment	D	14	3	82.4
		T	5	9	64.3
		Overall Percentage			74.2

a. The cut value is .500

  - During 10-fold cross validation accuracy falls to 64.5%
  - Deceptive 76.4%; Truthful 50.0%

Conclusions

- **Environmental issues severely hamper the applicability of kinesic analysis**
  - Surroundings
  - Equipment
  - Occlusion
- **For kinesic features to contribute to credibility assessment, environment must be carefully controlled**
  - No objects permitting occlusion or movement
  - Hi-quality cameras (HD cameras are fairly inexpensive)
  - Sufficient lighting

Conclusions

- **Blob analysis and ASM analysis yielded modest results**
  - Partially due to small sample size
  - Indicators with small effect sizes must be combined into models or indices (similar to kiosk deception index)
  - Caution must be exercised when interpreting the results
- **There may be some incremental value in capturing kinesics using blobs and ASMs, but there may be easier more diagnostic methods for doing it**

Conclusions

- **Blinking behavior provided initial evidence of diagnosticity**
  - Liars blink for longer periods in response to early, affect-related questions
  - Blink duration was diagnostic of deception and improved classification accuracy
- **Blinking behavior may be influenced by question effects**
  - Unclear if question effect is temporal or due to individual questions
- **Blinking behavior is inconsistent with other work**
  - May be difference between frequency and duration
  - Long blinks may be gaze aversion or eye closure

Conclusions

Questions or Comments?