Face Factors: Data reduction and Computer vision for Behavioral Science

Timothy R. Brick

Human Dynamics and Family Studies
Penn State University

Biostatistics, Epidemiology and Research Design
April 12, 2016
A List of Thanks

to my collaborators and colleagues

- Allison Gray and my PSU undergrads (PSU)
- Conrad Tucker and Shruthi Bezawada (PSU)
- Roland Göeke, Jason Saragih (U Canberra)
- Simon Lucey (Carnegie Mellon)
- Andreas Brandmaier (MPI-Bildungsforschung)
- Angela Staples (Eastern Michigan University)
- Michael D. Hunter (U. of Oklahoma)
- Barry-John Theobald (University of East Anglia)
- I. Dzobiek, N. Green (Humboldt Univ. Berlin)
Methods Background

PCA and SVD

**PCA**

\[ Z = X^T X \]
\[ Z = V D V^T \]
\[ L_X = X V \]

**Factor Analysis**

\[ Z = X^T X \]
\[ Z = A D A^T + e \]
\[ L = X A \]

- Decompose covariance into component parts
- PCA: Optimal representation
- Factor Analysis: “Sensible” representation
- Great for data reduction: We can choose how many components to keep
- Can be “rotated” to arbitrary arrangements

(e.g., Hotelling, 1933; Jöreskog, 1969; Krishnan, et. al. 2011)
Factor Analysis

Depression Measure 2 (Z-Score) vs. Depression Measure 1 (Z-Score)
Factor Analysis

![Factor Analysis Diagram]

- Depression Measure 1 (Z-Score)
- Depression Measure 2 (Z-Score)
Substantive Background

Conversation

(Image courtesy of www.business.com)
Mirroring in Conversation

- Mirroring Happens In Conversation
- Increased mirroring associated with increased rapport
- Mirroring appears to be non-conscious (until I mention it)

- What information is contained in this process?
- Something about rapport? Information?

- Seems to break down in communication disorders (e.g. Autism, schizophrenia)

(as far back as Cappella, et al., 1991; LaFrance, et al., 1985)
A Model of Mirroring

[Diagram showing the interactions between Cognition, Auditory Processing, Visual Processing, Motor Control, Mirror System, and Conversant A and Conversant B.]
Technology for Measurement
Old-fashioned Magnetic Motion Capture

Head Sensor
Chest Sensor
Arm Sensor
Hand Sensor
Knee Sensor
Background

History

Avatars

Emotion in Context

Current Directions

Conclusions

Technology for Measurement

Old-fashioned Magnetic Motion Capture

- Head Sensor
- Chest Sensor
- Arm Sensor
- Hand Sensor
- Knee Sensor
Technology for Measurement

Old-fashioned Magnetic Motion Capture

Head Sensor

Chest Sensor

Arm Sensor

Hand Sensor

Knee Sensor
Movement and Synchronization

Two talkers talking

Two dancers dancing
Movement and Synchronization

Two dancers dancing

Two talkers talking
Previous Findings

Yay!

- People (in general) are more synchronized in conversation.
- More conversational dominance means you lead more.
- But dominant people move less.

(e.g. Rotondo, et al., 2002)
Previous Findings

Yay!

- People (in general) are more synchronized in conversation.
- More conversational dominance means you lead more.
- But dominant people move less.
- **Women move more than men, regardless of dominance.**
- Everyone moves more talking to a woman than to a man.

(e.g. Rotondo, et al., 2002)
Previous Findings

Yay!

- People (in general) are more synchronized in conversation.
- More conversational dominance means you lead more.
- But dominant people move less.
- **Women move more than men, regardless of dominance.**
- **Everyone moves more talking to a woman than to a man.**
- Women synchronize with women better than with men.
- Men synchronize with women better than with men.

(e.g. Rotondo, et al., 2002)
Previous Findings

Yay!

- People (in general) are more synchronized in conversation.
- More conversational dominance means you lead more.
- But dominant people move less.
- **Women move more than men, regardless of dominance.**
- **Everyone moves more talking to a woman than to a man.**
- Women synchronize with women better than with men.
- Men synchronize with women better than with men.
- **Actually, men synchronize with people they aren’t talking to better than with men they are.**

(e.g. Rotondo, et al., 2002)
Methodology Note:

To get that last finding, we needed surrogate data analysis.

1. Get data from the same people
2. Generate “Psuedo-dyads” of people NOT talking to each other
3. Build a null distribution (1000s of reps)
4. Compare to real data

**Result:** synchronization for Male-Male conversations is LESS THAN “chance”.

But why?
Theoretical Implications

Aha!
Dynamic Equilibrium!

Stock photo courtesy of istockphoto.com. Non-commercial use permitted under their standard license.
Theoretical Implications

Aha!
Dynamic Equilibrium!

Aha!
Socialized Expectations!

Stock photo courtesy of istockphoto.com. Non-commercial use permitted under their standard license.
Representation

How do you model the face of a model?

\[ X = \begin{bmatrix}
  x_{11} & y_{11} & x_{21} & y_{21} & \cdots & x_{N1} & y_{N1} & P_{x1} & \cdots \\
  x_{12} & y_{12} & x_{22} & y_{22} & \cdots & x_{N2} & y_{N2} & P_{x2} & \cdots \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\
  x_{1m} & y_{1m} & x_{2m} & y_{2m} & \cdots & x_{Nm} & y_{Nm} & P_{x1} & \cdots \\
\end{bmatrix} \]
How Does It Work?

Active Appearance Models

\[ \hat{\mathbf{s}} = \mathbf{s}_0 + \mathbf{s}_a \mathbf{p}_s \]
\[ \hat{\mathbf{a}} = \mathbf{a}_0 + \mathbf{a}_a \mathbf{p}_a \]

(Cootes, 2001)
Evaluate the transformation
You can do this yourself!
Rotation in Face Space

\[ X^T X = WD W^T \]
\[ L_X = XW \]

(Theobald et. al., 2009; Brick, et. al., 2009b)
Rotation in Face Space

\[ X^T X = WDW^T \]
\[ L_X = XW \]
\[ \hat{Y} = L_X MV^T \]
\[ Y^T Y = VEV^T \]
\[ M_{ij} = \langle W_i, V_j \rangle \]

(Theobald et. al., 2009; Brick, et. al., 2009b)
Another Demo

Aren’t these fun?

(Brick, et. al., 2009; Boker et al., 2011)
The Model
What can we do with this model?
The Model

What can we do with this model?

![Diagram showing the model with components such as Cognition, Auditory Processing, Motor Control, Visual Processing, Vocal Processor, Avatar Processor, and Mirror System connecting Conversant A and Conversant B.]
### Findings from face-swapping

#### In numbers

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>SE</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>13.324</td>
<td>0.5261</td>
<td>(&lt; .0001)</td>
</tr>
<tr>
<td>Actor is Male</td>
<td>-2.769</td>
<td>0.2385</td>
<td>(&lt; .0001)</td>
</tr>
<tr>
<td>Partner is Male</td>
<td>-1.772</td>
<td>0.2606</td>
<td>(&lt; .0001)</td>
</tr>
<tr>
<td>Actor is RA</td>
<td>-0.872</td>
<td>0.1924</td>
<td>(&lt; .0001)</td>
</tr>
<tr>
<td>Avatar is Male</td>
<td>-0.070</td>
<td>0.1820</td>
<td>0.7016</td>
</tr>
<tr>
<td>Partner Vertical RMS</td>
<td>-0.484</td>
<td>0.0506</td>
<td>(&lt; .0001)</td>
</tr>
<tr>
<td>RA × Avatar Sex</td>
<td>0.008</td>
<td>0.3598</td>
<td>0.9814</td>
</tr>
<tr>
<td>RA × Partner Sex</td>
<td>-0.257</td>
<td>0.4720</td>
<td>0.5859</td>
</tr>
</tbody>
</table>

Observations= 310, Groups=28  
AIC=1043.2, BIC=1080.3  
Random Intercept SD=1.527, Residual=1.494

(e.g., Boker, et al., 2011)
Findings from face-swapping
In words

- Mixed-effects regression, grouped by participant, show:
  - Males move less
  - Everyone moves less when talking to a males
    - Regardless of who they THINK they’re talking to
  - If I move less for now, you move more (Turn-taking)
  - What matters to how you move is not who you THINK you’re talking to, but how the other person moves.

(Boker et al., 2011)
Moving Forward: Dynamic Manipulation

- So we changed static identity
- Can we modify dynamics?
  - Can we modify dynamics?
  - Well, Kinda.

(Cohn, et al, 2009)
Perturbing the system
Let’s try reduction

- Using RMS vertical and horizontal head velocity (captured by motion-tracking)
- Mixed-model regression, grouped by participant, show:
  - If I move less, you move more (Over any minute span)
  - But if one of us is damped, we BOTH move more

- This was supposed to emulate depression (but doesn’t)
- Maybe we’re not getting emotional data?
Hang on a sec...

This is actually an important question.

Before we can change emotions, let’s make sure we can first recognize them.

<table>
<thead>
<tr>
<th>AU4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brows lowered and drawn together</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AU5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper eyelids are raised</td>
</tr>
</tbody>
</table>
Coding Validation

Just the FACS

- The tracking representation captures a majority of the information of FACS codes
  - Tracking Points + Dynamics on Novel faces: mean overall classification rate was 90.2%
  - Manual FACS Coding: inter-rater agreement just over 93%
  - So we’re pretty close
- And the AAM representation can be manipulated and reconstructed

(Brick, Hunter, & Cohn, 2009)
Interesting Items About Emotion

1. Using clips from natural videoconference
2. 8-factor space
   ▶ Joy, Anger, Shame, Disguise, Sadness, Calm, Anxiety, Interest
3. An “interest” factor
   ▶ Surprise
   ▶ Interest
   ▶ Compassion
4. 47 out of 48 clips: significantly non-zero on more than one factor
   ▶ 33 out of 48 clips: more than three emotions
   ▶ Unclear whether this is concurrent or consecutive
Automatic Emotion Classification

(e.g. Bezawada, et al., submitted)
Non-emotional movements

Participant video
(identical and unaltered)

Research Assistant Video
(One simulated, one unaltered)
Avatar Movement Synthesis: Making Faces

But not at people

- Recorded AAM Data
- Recurrent Neural Network
  1. Input: Last 8 shape parameters from both participants plus sound level in that booth
  2. Output: Shape of RA in next frame
Expanding the representation

Newer models (called *multilinear models*) allow us to separate identity, emotional expression, and speech movements in a single model.

(Yin, et al., 2006; Bolkart & Wuhrer, 2015)
We suspect that synchronization is a measure of *rapport*.

- “Smoothness of conversation”
- Associated with:
  - Increased liking of interactor
  - Increased trust/belief
  - Increased agreement
- Success in person-centered (Rogerian) psychology
Getting into the real world

Fig 1. Video-based quantification of nonverbal behavior with motion energy analysis (MEA)

Fig 3. Correlation between synchrony and “quality of the therapeutic bond” from the therapist’s perspective.

Conclusions

- Our results show that nonverbal synchrony is closely related to symptom profiles.
- The amount of synchronized movement (grand mean of cross-correlations) was positively correlated with the patient’s evaluation of the “quality of the therapeutic bond.”

Discussion

- As the therapy progressed from sessions 1 to 40, so did the amount of pacing versus leading indicated that the imitation of the therapist, i.e., treatment pacing significantly higher in genuine interactions in this therapy setting.
- The therapy sessions compared to those of the patient’s predominant pattern were shuffled.

PLOS ONE | DOI:10.1371/journal.pone.0145882 December 30, 2015
Implications

I’m hoping to begin applications of these techniques soon:

- Training for Autism (with I. Dziobek, Humboldt University)
- Training for clinicians (Looking for Collaborators!)
- Better models for generation (deep learning?)
- Tracking of rapport/understanding for education and health communication
Summary

- Avatars & Machine learning can help us study interaction
  - Automatic analysis of facial movements
  - Near-perfectly controlled stimuli
  - Live interactive system perturbation
  - Inference about relationship/rapport (in process)
- And all because of a different way of applying factor analysis
Summary

- Avatars & Machine learning can help us study interaction
  - Automatic analysis of facial movements
  - Near-perfectly controlled stimuli
  - Live interactive system perturbation
  - Inference about relationship/rapport (in process)
- And all because of a different way of applying factor analysis

Thank You.

Timothy Brick <tbrick@psu.edu>