

Cultural Differences in Facial Expression Classification

Matthew N. Dailey

UCSD Computer Science & Engineering

Michael J. Lyons

ATR Media Information Science Labs

Miyuki Kamachi

ATR Human Information Science Labs

Hanae Ishi and Jiro Gyoba

Tohoku University Department of Psychology

Garrison W. Cottrell

UCSD Computer Science and Engineering

Introduction

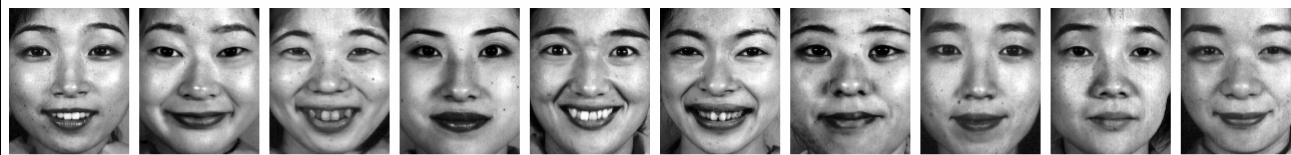
- There have been many cross-cultural studies of facial expression recognition but few attempts to understand what causes the observed differences between cultures.
- Here we interpret differences in terms of a neurocomputational model of visual perception and classification.
- We collected U.S. and Japanese ratings of the faces in the JAFFE database (Japanese Female Facial Expressions, Lyons, Akamatsu, Kamachi, and Gyoba, 1998).
- Behavioral results:
 - Japanese find many of the stimuli more intense (unlike in JACFEE studies).
 - Americans recognize the fear signal better.
 - Japanese recognize anger signals better, and attribute more disgust.
 - Asians in US are more like other Americans than like native Japanese.
- Our model suggests response bias as a major factor in the cultural differences.

Study of Facial Expressions

- Darwin (1872):
 - Found cross-cultural agreement on several emotional expressions.
 - Argued that facial expressions were adaptive, evolved, innate.
- Cultural relativists (early 20th century):
 - Argued emotion was completely culturally relative.
 - Main evidence: a lack of agreement on emotion in expressions.
 - Rejected Darwin's anecdotal evidence.
- Schlosberg (1952): a structural theory:
 - Expression confusion patterns are not random.
 - Analysis of "errors" leads to a circular similarity space.
 - Dimensions are pleasure/displeasure vs. attention/rejection.
- Tomkins, Ekman, and others (late '60s to present):
 - Certain (carefully selected) emotional expressions are universal.
- Russell and others ('80s to present):
 - Structural theories and fuzzy categories better explain the data.

JAFFE (Lyons et al., 1998)

Hap



Sad

(Maudlin)



Fear



Ang



Sur



Dis



Neu



1

2

3

4

5

6

7

8

9

10

First-ever
database of
native
Japanese
portraying
Ekman's*
basic
emotions.

** Not FACS
certified*

U.S. Ratings

- 99 subjects:
 - 39 Asians (including 8 Japanese), 34 Caucasian, 12 Hispanic, 8 Other/Mixed, 3 Middle Eastern, 2 Indian, 1 African.
 - 10 of the Asian subjects were raised in Asia, 29 in the U.S.
- Image data:
 - Cropped and scaled to size and positions used by our model.
 - Split randomly into two sets containing 5 models each.
 - Each subject viewed one set (110 or 107 images) in random order.
 - For each image and each emotion, an intensity rating from 1-5 was input with mouse and radio buttons.
- Subjects' sex, age, race, culture, and "amount of exposure to Asian culture" were recorded.

Japanese Ratings

- 30 Japanese undergraduates at Tohoku University.
- Same procedures for rating emotional intensities.
- Age, race, and sex were recorded.
- Subjects were also asked about prior exposure to Western people and culture:
 - # of Western movies seen: none (9), 1-4 (17), ≥ 5 (4).
 - Pop music style: Japanese (19), Am/Brit (9), none (2).
 - English ability: none (5), can understand, not converse (18), can have simple conversations (7), fluent (0).
 - Travel overseas: never (18), one trip (9), several trips (3).
 - Lived overseas: never (27), 1-3 months (1), 6-12 months (1), ≥ 1 year (1).
- Overall, little exposure to Western culture.

Asians in the US

- Are Asians in the US more like other US residents or more like the native Japanese subjects?
- We computed average responses of Japanese, non-Asian San Diego subjects, and Asian San Diego subjects.
- Pair-wise correlations between populations' responses:

	Japanese	US non-Asian	US Asian
Japanese	1.0		
US non-Asian	0.888	1.0	
US Asian	0.895	0.950	1.0

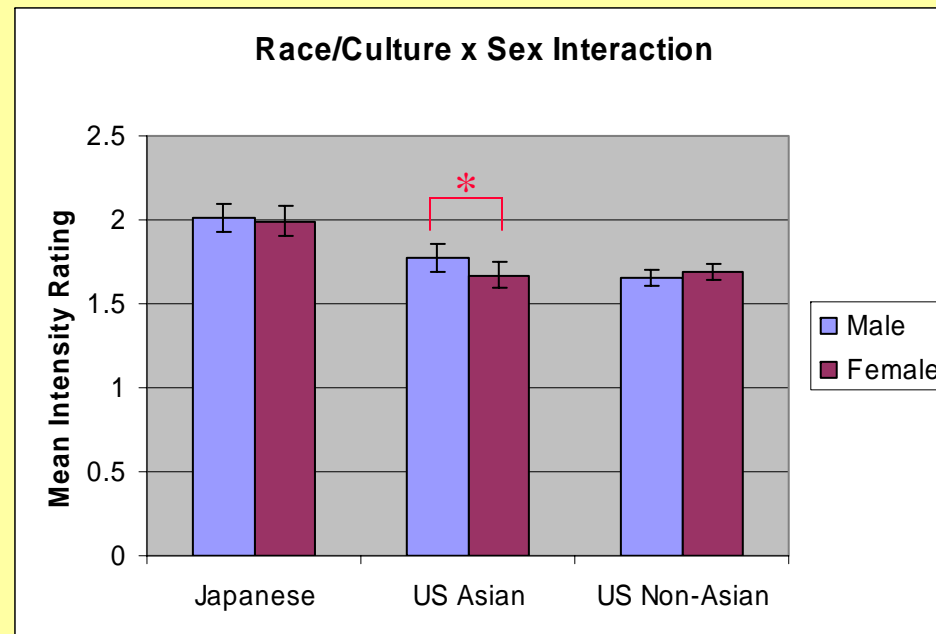
- US Asians are more like other US residents than like native Japanese.
- Viewer culture dominates viewer race.

Analysis of Intensity Ratings

- 4-way ANOVA on 7x6 confusion matrix computed for each subject.
- Factors:
 - Group: Japanese, U.S.-Asian, or U.S.-Non-Asian. 10 U.S. Asian-raised subjects were left out.
 - Subject Sex
 - Posed Emotion (H/M/F/A/S/D/N) (M for Maudlin)
 - Rated Emotion (H/M/F/A/S/D)
- Expected interactions:
 - Posed x Rated (confusion matrix entries differ)
 - Group x Posed (overall intensity rating of an expression differs by culture)
 - Group x Rated (overall attribution of an emotion differs by culture)
 - Group x Posed x Rated (some entries of confusion matrix differ by culture)
- Results ($\alpha_{FW} = 0.05$ over all possible interactions):
 - All four predicted effects were confirmed.
 - Unpredicted effect: Group x Sex interaction.
 - Within U.S. population alone, Race x Sex is significant, but there are no other significant interactions involving Race or Sex.

Post-hoc Analysis of Intensity

- The intensity rating analysis revealed an unpredicted interaction between Race/Cultural Group and Sex:



* = significant at $\alpha=0.05$ with Tukey's HSD

- The Japanese subjects attribute more emotion to the stimuli than do U.S. subjects (contra Matsumoto & Ekman)
- Asian men raised in the U.S. attribute more emotion to the stimuli than do Asian women or non-Asians.

Confusion Matrices

- How intense do the expressions look, and which expressions seem similar?

	H	M	F	A	S	D
H	4.00	1.19	1.16	1.17	1.36	1.17
M	1.22	3.60	2.03	1.83	1.23	2.37
F	1.13	2.70	3.23	1.84	2.78	2.96
A	1.13	1.93	1.44	4.03	1.22	2.89
S	1.60	1.34	1.94	1.35	4.36	1.45
D	1.10	2.44	1.99	2.62	1.39	3.94
N	1.55	1.54	1.30	1.68	1.32	1.51

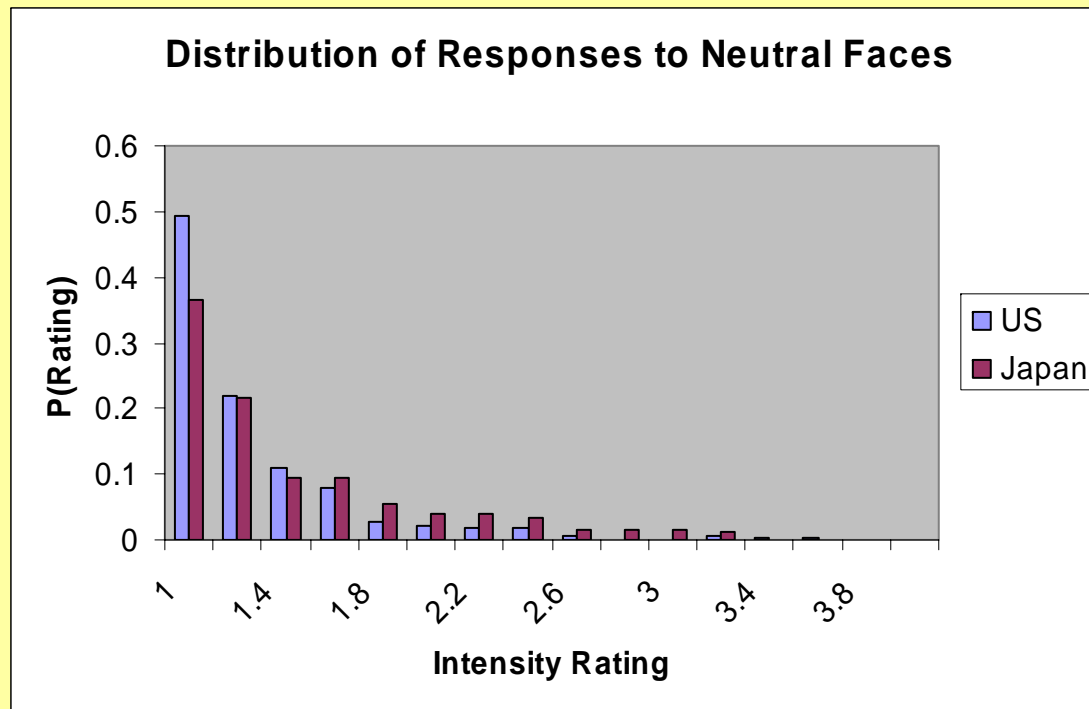
Japan (mean 2.00)

	H	M	F	A	S	D
H	4.10	1.07	1.04	1.04	1.32	1.04
M	1.16	3.17	1.59	1.66	1.08	1.52
F	1.04	1.91	3.05	1.38	2.03	1.75
A	1.04	1.94	1.16	3.08	1.06	1.94
S	1.48	1.08	2.03	1.10	4.13	1.16
D	1.03	2.15	1.44	2.32	1.22	2.91
N	1.41	1.49	1.17	1.41	1.16	1.18

U.S. (mean 1.69)

Intensity Bias

- A pilot study predicted Japanese subjects would find the neutral faces more intense.



- We created histograms from the subjects' average ratings to neutral faces in each population.

- The neutral-face distributions are significantly different (KS=0.147, $p=0.0054$)

Cultural Differences in Intensity

- 42 one way F tests ($\alpha_{FW} = 0.05$) revealed significant differences.
- 7 row-wise and 6 column-wise F tests ($\alpha_{FW} = 0.05$) also revealed significant differences.

		Response Emotion						
		H	M	F	A	S	D	
Stimulus Emotion	H		J				J	
	M		J				J	J
	F		J		J	J	J	J
	A				J		J	J
	S		J		J		J	
	D		J				J	J
	N						J	
					J	J	J	

- Japanese attribute more emotion to negative faces.
- Japanese attribute more anger, surprise, and disgust.
- Matsumoto and Ekman (1989): opposite trend (Japanese subjects rated JACFEE faces as *less* intense)

Relative Perceived Intensity

- Are relative levels of intensity different across cultures?
- We averaged each subject's intensity ratings across faces showing the same emotion.
- Within each population (Japanese, US-Asians, and Non-Asians in the US) we compared mean rated intensities of the portrayed emotion with a one-way ANOVA and Tukey's correction for multiple pair-wise comparisons.
 - Japanese: $S=H=D=M=A=F$ (H>M, D>A)
 - US Asians: $S=H>M=A=F=D$
 - Non-Asians in US: $S=H>M=A=F=D$
- Matsumoto and Ekman did same comparison with JACFEE:
 - Japanese: $D>H>A>S>M$
 - Non-Asians in US: $H=A>D>S>M$
- As with JACFEE, Japanese perceive disgust relatively more strongly.

Accuracy

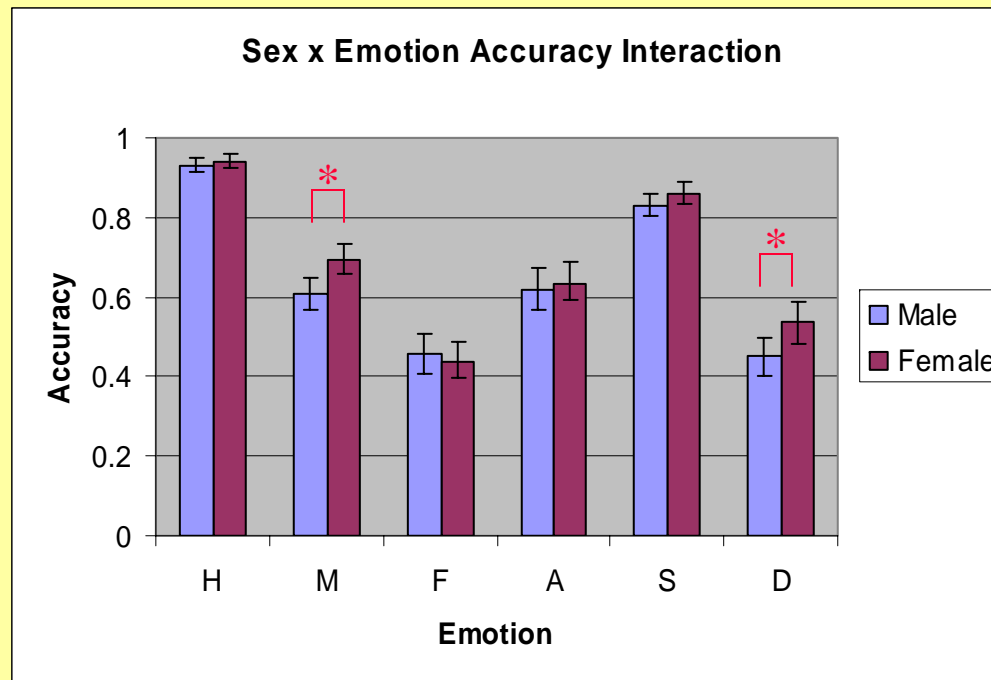
- “Accuracy” is % agreement on *intended* emotion.
- But subjects are free to assign tied top ratings.
- Japanese subjects do this more often: 25% vs. 17%.
- When top rating is tied, assign partial credit for the tie:
 - Credit = $1/n$, where n is the number of maximally-rated emotions.
- 3-way ANOVA (Group, Sex, Posed-Emotion)
- Results ($\alpha_{FW} = 0.05$ over all possible interactions):
 - Confirmed predicted effect: Group x Emotion interaction.
 - Unpredicted: Sex x Emotion interaction; Sex marginal effect.

	H	M	F	A	S	D	Overall
Japan	0.921	0.653	0.297	0.716	0.883	0.636	0.684
US	0.942	0.656	0.495	0.598	0.833	0.452	0.663

■ =
significant
($\alpha_{FW}=0.05$)

Post-hoc Analysis of Accuracy

- Accuracy analysis revealed a significant interaction between Sex and Posed-Emotion in accuracy.



* = significant at $\alpha=0.05$ with Tukey's HSD

- Female subjects are better at detecting sadness and disgust.

Relative Recognition Difficulty

- Within each culture (Japanese and Western), we compared mean accuracy on each emotion using Tukey's HSD.
- JACFEE analysis is from Matsumoto and Ekman (6-way forced choice).

	JAFFE	JACFEE
U.S. Asian	H = S > M = A > F = D	
U.S. Non-Asian	H > S > M = A > F = D	H > M = S = D = A > F
Japanese	H = S > A = M = D > F	H > S > D = M > A > F

- JACFEE ratings are relatively consistent across cultures.
- JAFFE ratings: disgust is recognized as easily as anger and sadness for Japanese; as hard as fear for U.S. subjects.

Sensitivity

- Do differences in accuracy reflect response bias or discrimination?
- d' is one way to measure raw discrimination ability.
Treating each emotion independently:

$$d' = \frac{\bar{p} - \bar{n}}{\sigma_n}$$

	H	M	F	A	S	D
Japan	8.45	2.33	2.76	3.47	4.10	2.07
US	8.45	2.74	3.48	2.73	6.34	2.88

- US subjects' accuracy on fear is due to improved detection.
- Japanese subjects' accuracy on anger is due to improved detection.
- Japanese subjects' accuracy on disgust is due to response bias.

Discussion of Behavioral Data

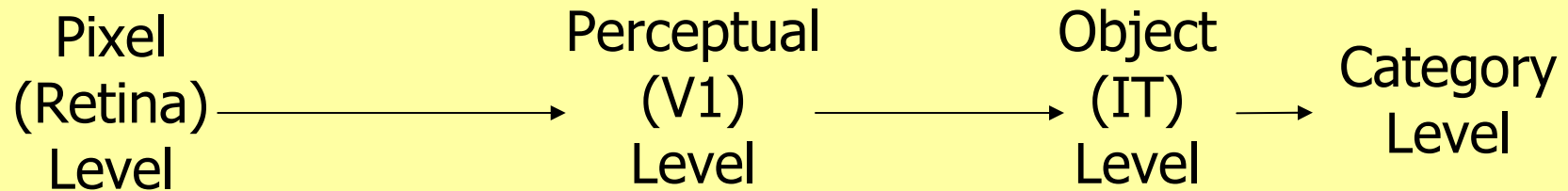
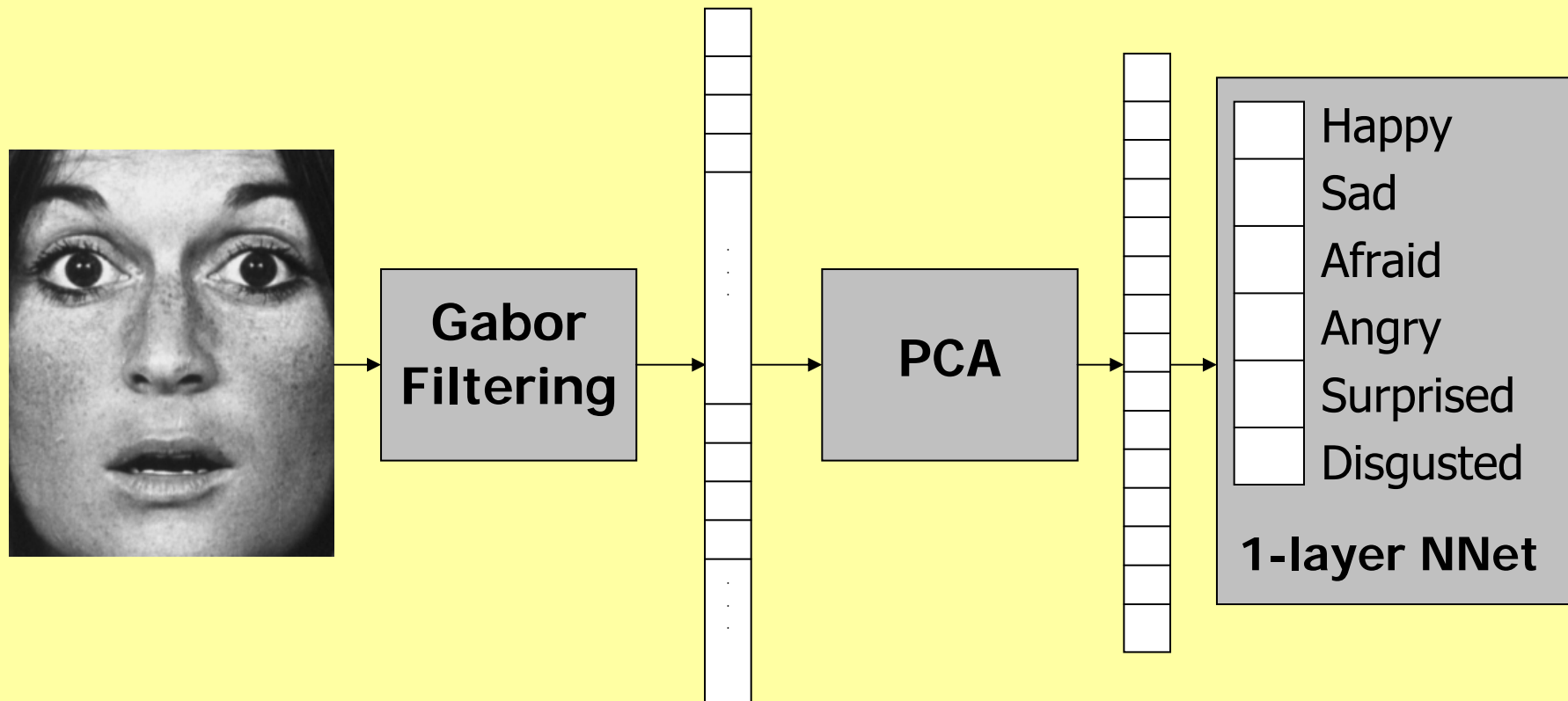
- We found no evidence of an other-race effect.
 - US Asians are no better than non-Asians at the task.
 - But the Race x Sex interactions deserve further study.
- We found interesting cross-cultural effects.
 - Japanese find many of the stimuli more intense, in contrast to previous results (Matsumoto & Ekman, 1989)
 - Americans recognize the fear signal better.
 - Japanese recognize the anger signal better.
 - Japanese may have a stronger response bias for disgust.
- Higher intensity ratings and improved accuracy on anger and disgust by Japanese suggests JAFFE is more consistent with Japanese experience (“training set,” for our model).
- Can we model cultural differences computationally?

Modeling Cultural Effects

- Japanese subjects are worse at recognizing fear than Americans, in both JACFEE and JAFFE.
- Some of the posed expressions in JAFFE appear culture-specific. E.g. puffing out the cheek in anger.
- Regardless of whether facial expressions are universal, expression/emotion mappings probably must be learned.
- Hypothesis 1:
 - Display rules and culture-specific forms prevent certain prototypical expressions from being learned well.
 - Perhaps JAFFE represents more natural expressions seen in Japan.
 - Train "Japanese" network on JAFFE. Train "US" network on POFA.
 - Test both on JAFFE and compare to human intensity ratings.
 - Prediction: the "Japanese" network's responses should correlate with Japanese human subjects, and similarly for "US" nets.

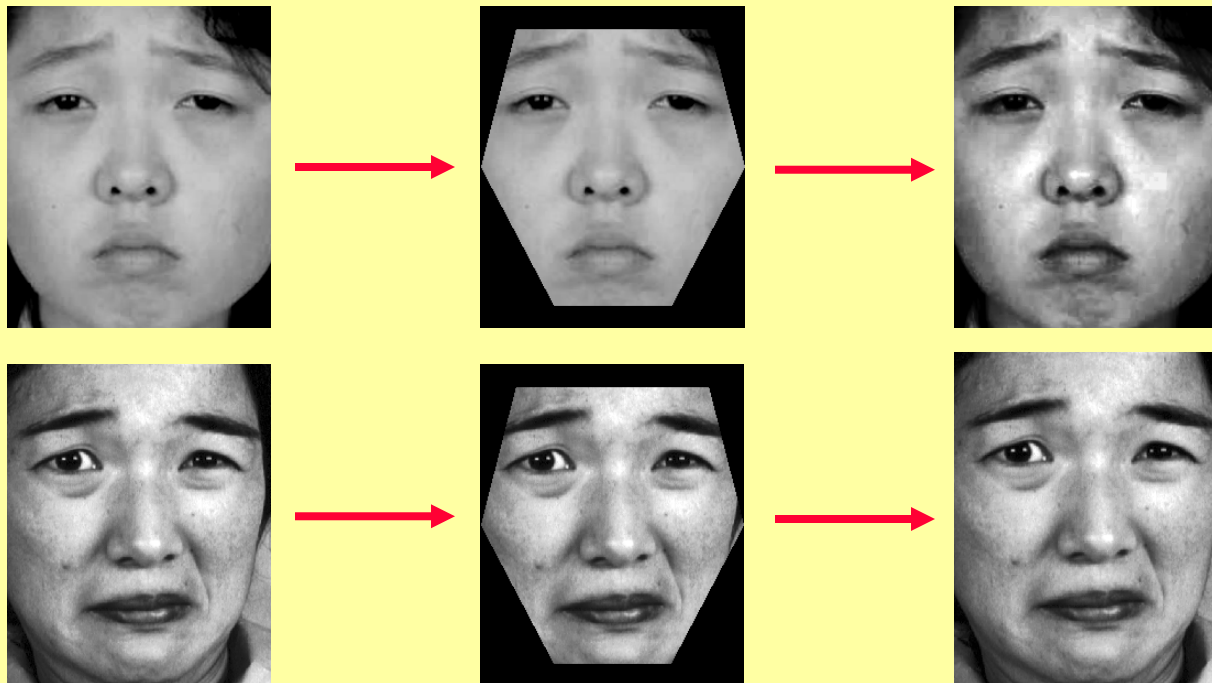
The Facial Expression Classifier

Dailey, Cottrell, Padgett, and Adolphs (2002)



Preprocessing

- In the databases (POFA and JAFFE), quality, lighting conditions, and background darkness vary.
- Simple average brightness and contrast normalization does not perform well.
- Solution: mask off background and hair and normalize histogram of unmasked region to the average histogram:

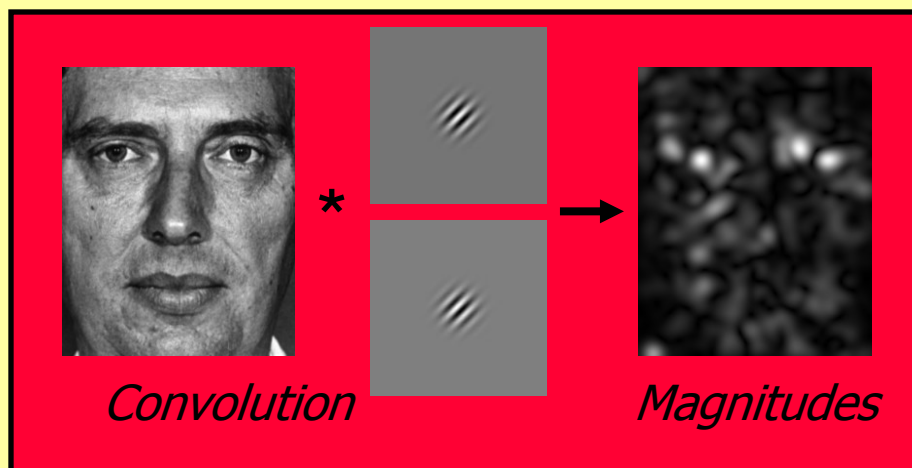


The Gabor Lattice Representation

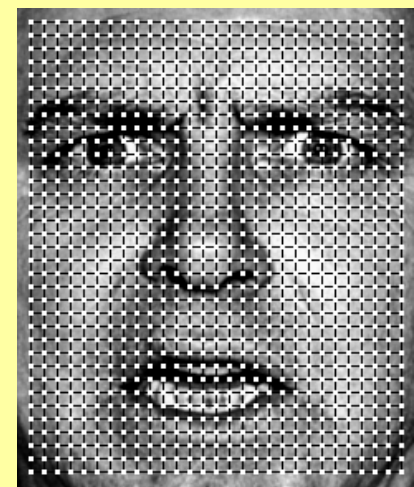
- Basic feature: the 2-D Gabor wavelet filter (Daugman, 85):



- Combine two filters to get phase insensitivity, modeling complex cell responses in primary visual cortex.

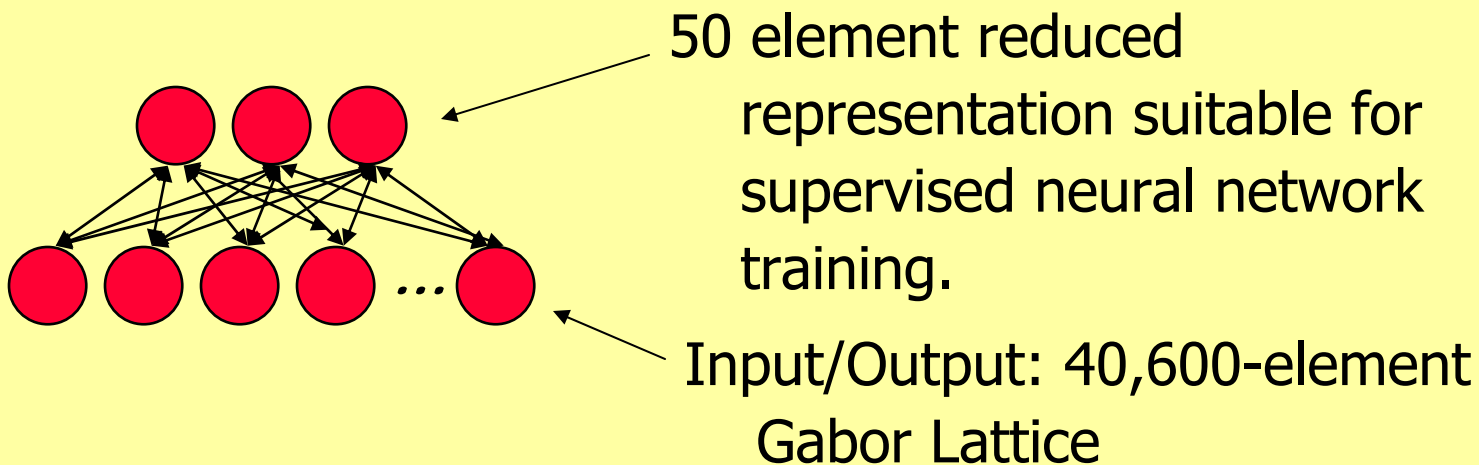


- Subsample in a 29x36 grid:



Principal Components Analysis

- Principal Components Analysis (PCA) for unsupervised dimensionality reduction:

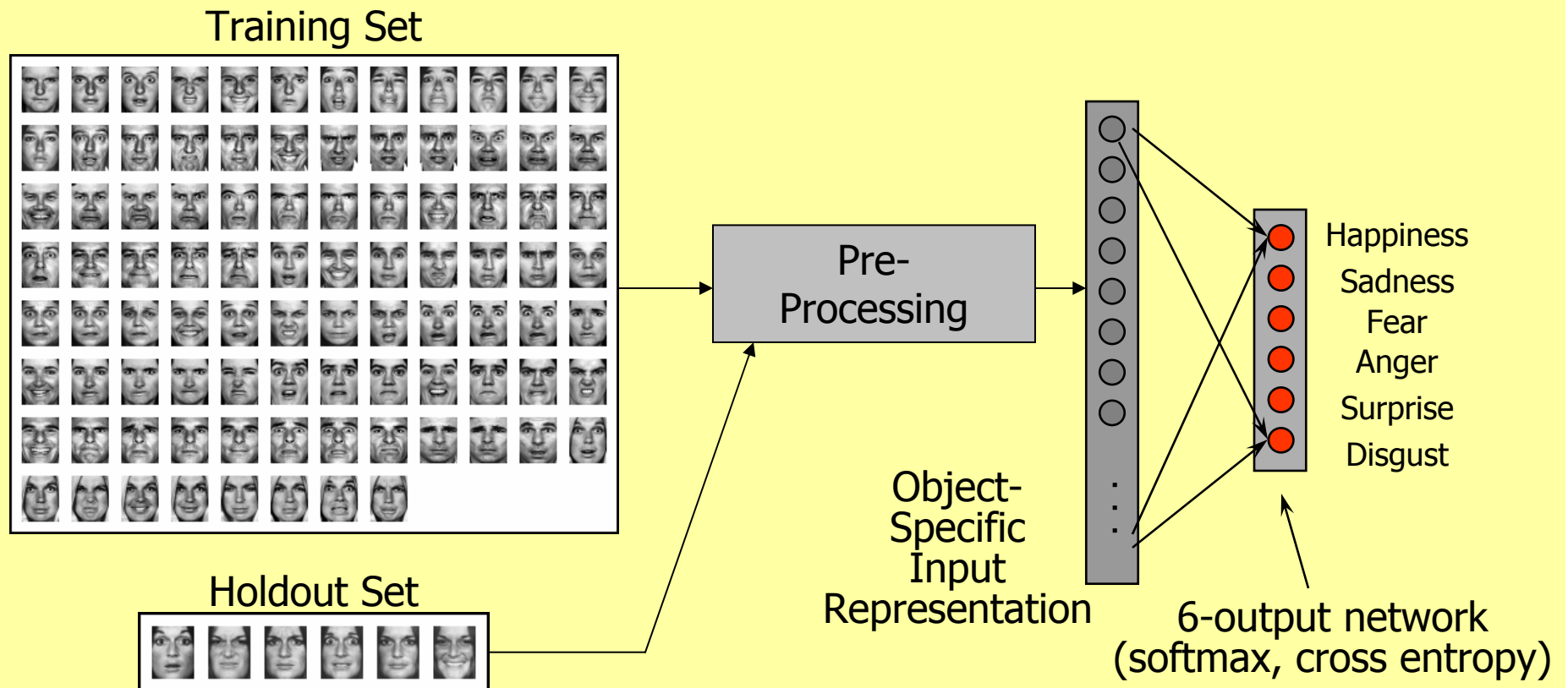


(PCA can be learned by a Hebbian network)

- The resulting 50 inputs are fed to the category layer: a 6-unit softmax network (no hidden layer).

The Final Layer: Classification

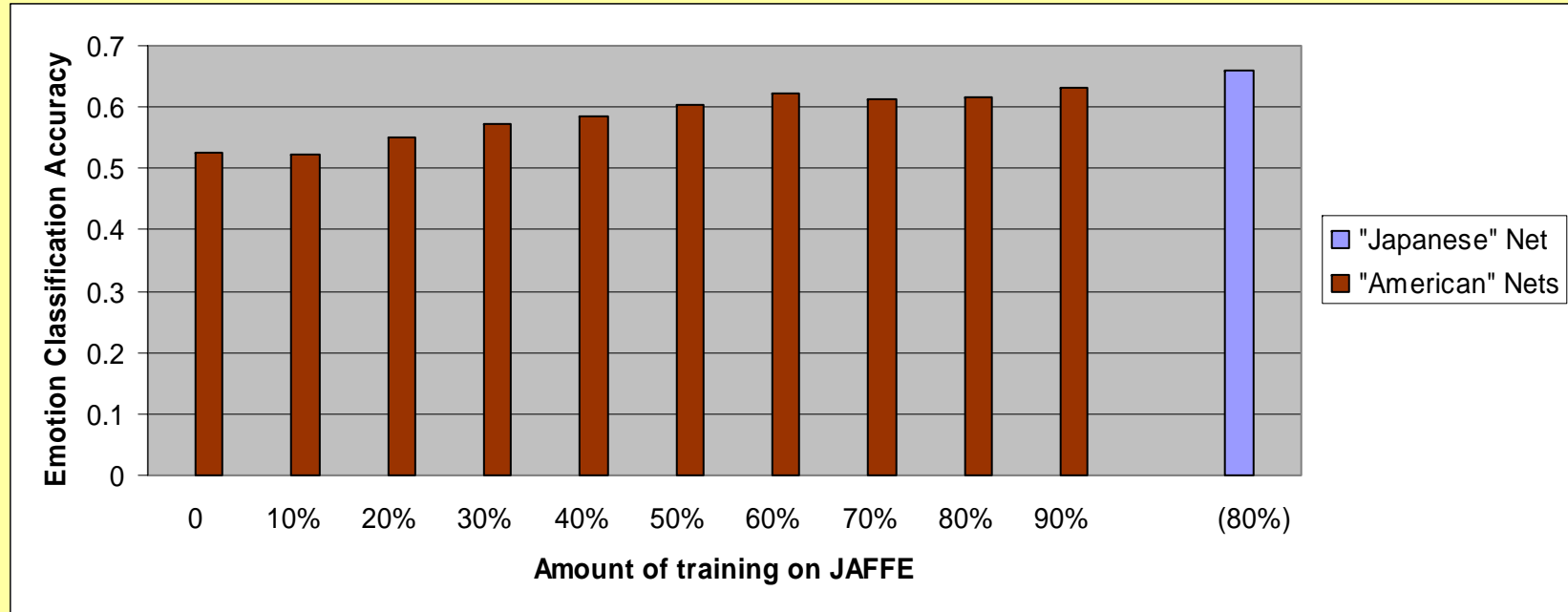
- Train the network to produce binary outputs for expressive faces and uniform vector for neutral faces.
- Stop training when holdout set performance is best.
- Test the network's generalization on images of a previously unseen actor's faces.



Methods

- “Japanese” networks:
 - PCA trained on 8/10 JAFFE models.
 - Each network is trained on the 8 JAFFE actors in PCA training set.
 - 9th model is held out to determine when to stop training.
 - Each network is tested on remaining 10th JAFFE model.
 - Cross-validated (90 different networks, 90 different PCA’s)
- “American” networks:
 - PCA trained on 13/14 POFA actors.
 - Each network is trained on those 13 POFA models.
 - 14th actor is held out to determine when to stop training.
 - Training exposure to JAFFE is a free parameter varying from 0/10 to 9/10 models.
 - Cross-validated (140 different PCA’s; 140 different networks at each training level – 1260 networks total).

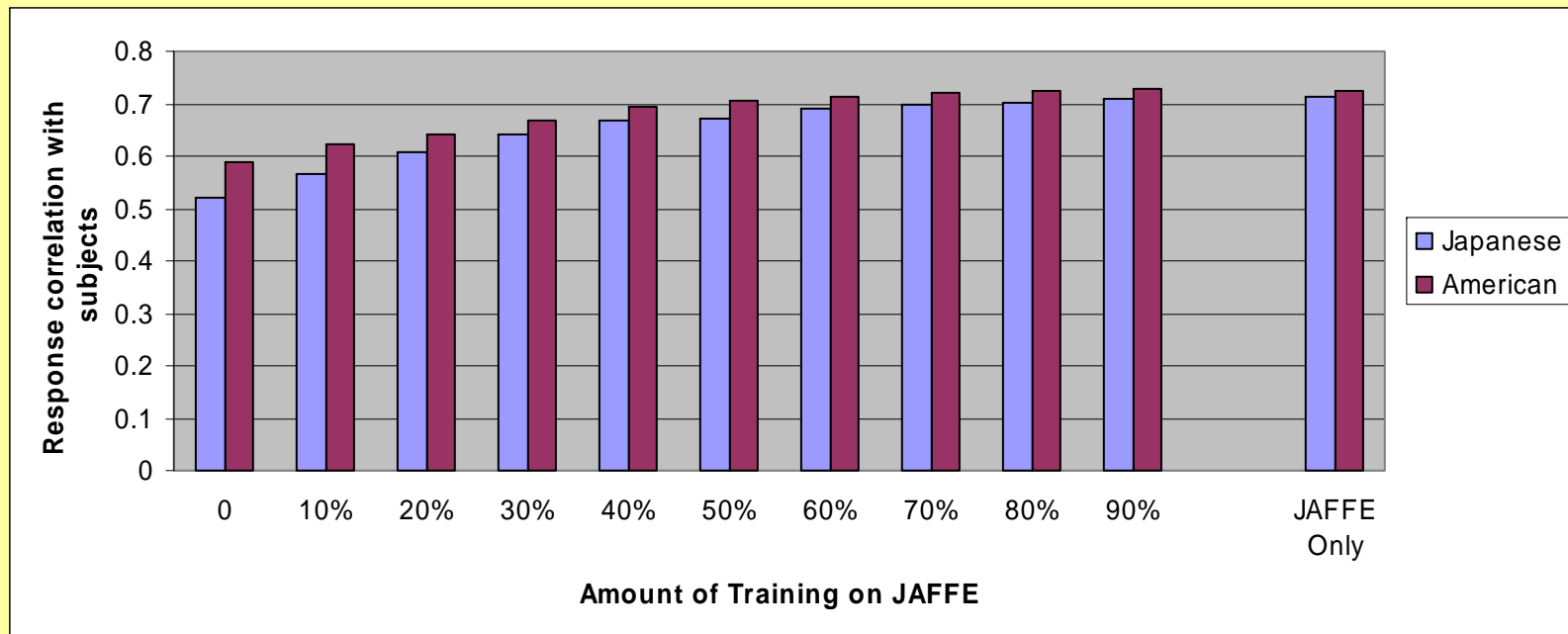
Network Accuracy Results



- Networks trained and tested on JAFFE alone perform best.
- Training on POFA apparently adds noise with respect to the JAFFE classification task.

Results – Correlation with Humans

- We first average the nets' 6-element responses to each emotional face.
- For each stimulus, we correlate the network and human responses.



- All the networks' responses look more American than Japanese.
- We must conclude that the difference between JAFFE and POFA as training sets does not explain cross-cultural differences in behavior.

Alternative Explanation: Response Bias

- Hypothesis 1 was incorrect: training on JAFFE does not lead to “Japanese” responses and training on POFA does not produce “American” responses.
- But training on JAFFE does produce good correlations with both populations.
- Is it possible to produce “cultural differentiation” among networks with the same training set?
- Hypothesis 2: response bias is one factor that can play a role. Japanese marginal anger, surprise and disgust responses were significantly higher than Americans’.
- Bayes’ rule tells us: $P(\text{emo}|\text{expr}) = P(\text{expr}|\text{emo})P(\text{emo})/P(\text{expr})$.
- Experiment: assume $P(\text{expr})$ is uniform for subjects and networks, and optimize priors $P(\text{emo})$ for networks to maximize correlation with American or Japanese subjects’ intensity judgments.
- The best-performing set of networks (trained on JAFFE only) was used for this experiment.

Response Bias Results

- Optimal priors $P(\text{emo})$ and resulting correlations:

	H	M	F	A	S	D	Japan corr.	US corr.
Japan	0.119	0.248	0.106	0.104	0.194	0.229	0.742	0.708
US	0.129	0.261	0.159	0.154	0.199	0.099	0.706	0.742

- Main difference: Japanese bias toward disgust.
- Per-emotion recognition accuracy for biased nets:

	H	M	F	A	S	D	Overall
Japan	0.752	0.538	0.340	0.511	0.856	0.682	0.613
US	0.778	0.631	0.545	0.663	0.815	0.356	0.633

■ =
significant
($\alpha_{FW}=0.05$)

- Differences occur on the same emotions as in human data, but for Anger, the difference is in the opposite direction!

Discussion

- Surprisingly, simply exposing an emotionless machine to the JAFFE expressions accounts for a large proportion of the variance of human subjects' responses.
- Also surprisingly, the composition of the training set may be a less important predictor of cultural differences than we first thought.
- Response bias in a network trained on JAFFE alone can account for some of the cultural variation in our study.
- The biggest gap between the bias model and humans was on the JAFFE anger expressions – the “American” network performs too well. This suggests the JAFFE anger expressions are legitimate but culture-specific.
- There is a lot of room for further research:
 - Correlation between American and Japanese subjects is $r=0.912$, but the best networks only achieve $r=0.742$.
 - There is no mechanism in the model to account for intensity bias.

References

- Dailey, M.N., Cottrell, G.W., Padgett, C., and Adolphs, R.A. (2002). A neural network that categorizes facial expressions. *J. Cognitive Neuroscience*. In press.
- Darwin, C. (1872/1998). *The Expression of Emotions in Man and Animals*. Oxford University Press, New York, 3rd edition.
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Optical Soc. of Am. A*, 2:1160–1169.
- Ekman, P. (1999). Facial expressions. In *Handbook of Cognition and Emotion*, pp. 45–60. Wiley, New York.
- Ekman, P. and Friesen, W. (1976). *Pictures of Facial Affect*. Consulting Psychologists Press, Palo Alto, CA.
- Landis, C. (1924). Studies of emotional reactions: II. General behavior and facial expression. *J. Comparative Psychology*, 4:447–509.
- Lyons, M.J., Akamatsu, S., Kamachi, M., and Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. *Proc. 3rd IEEE Conf. On Automatic Face and Gesture Recognition*, pp. 200–205.
- Matsumoto, D. and Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion*, 13(2):143–157.
- Russell, J.A. and Bullock, M. (1986). Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition*, 4(3):309–341.
- Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology*, 44:229–237.
- Tomkins, S.S. (1962–1963). *Affect, Imagery, Consciousness*, vol. 1–2. Springer, New York.