Is Facial Expression Processing Holistic?

Akinyinka O. Omigbodun (aomigbodun@ucsd.edu)
Electrical and Computer Engineering, University of California San Diego
La Jolla, CA 92093 USA

Garrison W. Cottrell (gary@eng.ucsd.edu)
Computer Science and Engineering, University of California San Diego
La Jolla, CA 92093 USA

Abstract

Most studies examine holistic processing with respect to facial identity, but at least one study has also looked at holistic processing of facial expressions (Calder, Young, Keane, & Dean, 2000). However, this work used the partial composite paradigm, which is known to exhibit bias effects (Richler, Cheung, & Gauthier, 2011). The complete composite paradigm (Gauthier & Bukach, 2007) provides a bias-proof way to quantitatively measure holistic processing. In this paper, we perform the corresponding experiment in our face processing model (EMPATH, (Dailey, Cottrell, Padgett, & Adolphs, 2002)) to predict whether holistic processing of facial expressions will be found if the corresponding human experiments are performed, and our prediction is that it will. We also compared our model's performance to the participants in recent experiments in facial expression recognition in humans (Tanaka, Kaiser, Butler, & Le Grand, 2012). Tanaka et al. (2012) concluded that expression recognition is not always holistic, but our results suggest that it is.

Keywords: holistic processing

Introduction

Is *all* facial processing holistic? For example, when we are judging whether a person is trustworthy, tired, middle-aged, or happy, is our decision based on a global percept of the face rather than constituent parts? One might suspect that the answer is no, at least in some cases. Holistic processing might depend upon the task. The evidence from face inversion, composite, and parts-wholes experiments weighs in favor of holistic over parts-based processing for face identity (Yin, 1969; Tanaka & Farah, 1993; Cheung, Richler, Palmeri, & Gauthier, 2008; Richler, Tanaka, Brown, & Gauthier, 2008). Less research has focused on understanding the nature of facial expression recognition. One reason to suspect that expression recognition is different is the categorization of expressions into just seven categories – happy, sad, surprised, angry, disgusted, fearful, and neutral – in line with the "six basic expressions" theory (Ekman & Friesen, 1976), at least as practiced in most psychology experiments. Previous work in our lab has shown that the basic level processing of objects (e.g., into an overall category, such as chairs or lamps), does not behave in the same manner as subordinate level processing in our models, and does not engage our model's equivalent of the Fusiform Face Area. One consideration is that the holistic processing of faces may be induced if there is any variation in identity, regardless of the task being performed. In our modeling work, we tested the hypothesis that without any variation in identity, the processing of facial expressions is holistic.

Holistic processing of facial stimuli is generally measured with composite face paradigms, where chimeric faces are constructed from the top and bottom halves of different "source" faces. Subjects are asked to identify face halves of chimeric faces or to judge whether two halves of a pair of chimeric faces are the same or different. Misalignment of the top and bottom halves generally leads to an increase in the subjects' accuracy and/or a decrease in reaction time relative to the aligned condition, demonstrating that faces are processed holistically. We test our model with simulations based on the complete composite paradigm (Gauthier & Bukach, 2007), an improvement upon what Gauthier & Bukach call the partial design, which is what is classically used (Young, Hellawell, & Hay, 1987). The complete composite paradigm eliminates the effects of response bias (for example, a preference towards answering "same" for misaligned stimuli). our results predict that a facial expression recognition experiment with humans based on the complete composite paradigm will indicate holistic processing.

We then use our model to account for experimental results in Tanaka et al. (2012). Tanaka et al. used their own nonstandard composite paradigm in one of their experiments that we simulated and a variation on the partial design in another that we also simulated. They inferred from their experimental results that facial expression processing is holistic when there is a clash between parts of a facial expression (e.g. angry-happy composite) but analytic or parts-based when there is little or no conflict between the parts (e.g. normal happy face). They posit an earlier stage where a stimulus is rapidly assessed for parts that clash to determine the processing pathway to be used. However, we achieved similar results with our model which (1) has one processing pathway for all stimuli, (2) does not have an earlier stage for quick appraisals, and (3) is not trained to combine decisions on parts of a stimulus into an overall decision. Our results suggest that one holistic processing pathway is sufficient to account for their data.

Complete Composite Paradigm (CCP)

In composite paradigms, participants in experiments are presented with two composite faces, and asked to make same-different judgments about the cued face halves (top or bottom) while ignoring the other face halves. The complete composite paradigm (CCP) (Gauthier & Bukach, 2007) provides a bias-proof way to quantitatively measure the interaction between a subject's decision and the presence or ab-

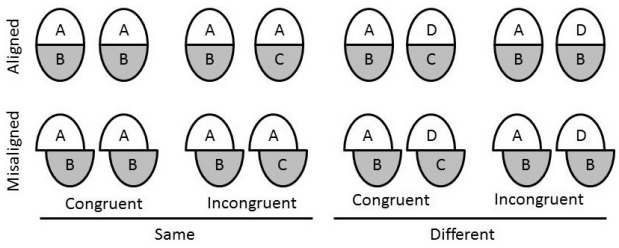

Figure 1: Schematic showing the complete composite paradigm. The subject must decide whether the top face halves are the same or different. In the congruent condition, the top and bottom face halves would generate the same answer, while in the incongruent condition, they would not. Holistic processing is measured as a congruency effect.

sence of a change in the unattended face halves (see Fig. 1). If the answer is the same for the top and bottom halves, it is a congruent trial, otherwise, it is incongruent. A congruency effect, difference in sensitivity, d', between congruent and incongruent trials, is indicative of holistic processing: the unattended face half is obligatorily processed. The partial design only has trials where the unattended halves are different, and holistic processing is measured as a difference in performance between aligned and misaligned trials. This is a flawed measure; using it, researchers concluded holistic processing was unrelated to performance on the Cambridge Face Memory Test (CFMT). The relationship is found to be strong using the complete design (Richler et al., 2011).

Methods

The Model

Each input image goes through a two-step preprocessing stage: Gabor filtering and Principal Component Analysis (PCA). The biologically motivated 2-D Gabor filter is constructed by using a two-dimensional sinusoid localized by a Gaussian envelope (Daugman, 1985). By tuning to particular spatial frequency and orientations, the Gabor filter magnitudes can be used to simulate the responses of complex cells in primary visual cortex (V1). Following Gabor filtering, PCA reduces the dimensionality of the information, simulating the information extraction mechanism beyond V1. After these preprocessing steps, each image is represented by a vector with relatively low dimension to be input to the perceptron.

We computed each Gabor filter using the following equation:

$$G(x,y,S,F,W) = k \cdot e^{-\pi S^2(x^2+y^2)} \cdot \left(e^{j(2\pi F(x\cos W + y\sin W))} - e^{-\pi(\frac{F}{S})^2}\right)$$

where S is a scaling parameter, (F,W) is the polar frequency of the complex sinusoid, (x,y) are the spatial coordinates and k is a constant (Dias, 2007; Movellan, 2002). There were five spatial frequencies ($F = 1/2^i$ for $i = 2,\ldots,6$) and eight spatial orientations ($W = j\pi/8$ for $j = 0,\ldots,7$) for a total of 40 different Gabor filters. S was related to F as follows: $S = 3F/\sqrt{2\pi}$. S was chosen such that each filter had the characteristic form of biological two-dimensional receptive field profiles (Hubel, 1988; Daugman, 1988). The filters were centered and applied at the 1080 points in a 36 by 30 grid on each image.

We then use PCA to map the Gabor filter features from a 43200-dimensional space to a space with many fewer dimensions. In forming the PCA projection matrix, we retained enough eigenvectors to account for 90% of the variance. For classification, a single-layer perceptron was trained with gradient descent using a cross-entropy error function.

Network Training and Testing

70 gray-scale images were selected from the Pictures of Facial Affect (POFA) (Ekman & Friesen, 1976). These were cropped to 240×292 pixels and adjusted to ensure that there was uniformity in the upright frontal face views; to allow for shifting, the input images were 1.5 times as wide as the face images (i.e. images of size 360×292) with the faces flush against the left or right edge, as shown in Fig. 3a. Misaligned face halves are shown in Fig. 3b.

The POFA dataset includes seven facial expressions (Happy, Sad, Surprised, Angry, Disgusted, Fearful, and Neutral) for each of 14 actors. We selected 10 of these for our experiments (em, gs, jj, mf, mo, nr, pe, pf, sw, wf). We trained on 9 and tested on 1, repeating this 10 times with each actor getting a chance to be tested. For the remaining 9 actors, we trained on 8 and used the 9th as a hold-out to stop training. This was repeated 9 times with each of the 9 actors having a turn as the hold-out; so we ended up with 9 networks predicting the expression on each stimulus. The consensus of the 9 instances was taken as the model's decision for an experimental run. We averaged the results of the 10 runs.

Experimental tasks in our simulations involving the identification of top and bottom face halves necessitated training the model to classify vectors of Gabor features corresponding to images with only half of a face. We modeled attention to half a face by using a transformation of the Gabor features that reduced the contribution of the top or bottom of a face to the total training set covariance. The transformation that we used for this and a proof that the contribution of transformed Gabor features to the total covariance is reduced is beyond the scope of this paper. In modeling the process of identifying top and bottom face halves in testing and in the experiments, we simulated giving more attention to half of a face using this transformation.

The PCA projection matrix was generated from the Gabor feature vectors designated for training the model. Both before and after the projection of the training stimuli by the matrix, the projections were z-scored in order to put each input to the perceptron on the same scale (LeCun, Bottou, Orr, & Mueller, 1998). The stimuli for cross-validation were rescaled as one set of vectors, and the stimuli for testing and the experiments were rescaled as another set; with the new sample variances for these sets equal to the ratio of the set sizes to the size of

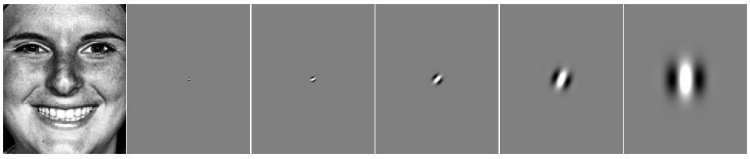

Figure 2: The real components of Gabor filters are shown relative to the size of an image at five scales and orientations.

a b

Figure 3: (a) left- and right-shifted images and (b) top-left-bottom-right (TLBR) and TRBL images

the training set. This was done to ensure that the rescaling was as uniform as possible across all the feature vectors.

All composites were constructed from the same individual (ten times, once for each test individual, as described above), which ensured that expression and identity recognition were not confounded. The networks were trained as described above, to classify each (whole, unaltered) face into one of seven expression categories. In order to obtain a same/different judgment from a network that only processed one face at a time, we presented the network with the two stimuli (with attention on the queried half, as described above), one at a time. Since there were actually nine networks for each test face (as described above), we compared the consensus of the networks on one stimulus with the consensus on the other. If they match, the overall response is "same"; otherwise the response is "different".

To obtain reaction times, we appeal to the well-established inverse relationship between reaction time and confidence ratings (Audley, 1960; Baranski & Petrusic, 1998) in mind, we computed a measure of confidence for our model which was the ratio of the number of network instances in agreement with the consensus to the total number of instances for each decision made (when the consensus was the correct decision). To model reaction time, we subtracted these values from one.

For the experiments by Tanaka et al. (2012) (described shortly), since there are no comparisons between two images (all the subjects had to do was decide if the cued half of the stimulus was happy or angry), we simply use a consensus of the nine networks.

Simulations

Simulation 1: The CCP on expression recognition The first simulation used the CCP (Fig. 1) to test whether our model predicts that facial expression recognition is holistic. Same-different judgments were obtained (as described above), and holistic processing was then measured as the difference in sensitivity, d', between congruent and incongruent trials.

Simulation 2: Experiment 1 of Tanaka et al. (2012) In experiments performed by (Tanaka et al., 2012), subjects were asked to decide if the cued half of the stimulus was happy or angry. Tanaka et al. (2012) measured the subjects' percentage accuracies and reaction times (in milliseconds). They based their experimental design on the observation that happy expressions are bottom biased (meaning it is easier to recognize a happy expression from the bottom of a happy face than it is from the top half) and angry expressions are top biased (Calder et al., 2000); they only counted trials in which the correct responses for the cued lower halves were happy, and those for the cued upper halves were angry. In their first experiment, Tanaka et al. (2012) compared subjects' performance for a happy lower half face in four pairings: (1) with a happy top half face (normal), (2) with an angry top half face (angry-happy), (3) without a top half face (isolated), and (4) with a neutral top half face (neutral).[1] There were four corresponding pairings for an angry top half face (see Fig. 4). Their reasoning was that if happy expressions in the lower face half and/or angry expressions in the top face half of normal, neutral, and isolated stimuli were recognized equally easily, then this was evidence for parts-based processing (since this would suggest that the other face half is being "ignored"); and if recognition of angry-happy stimuli was worse than that of neutral and isolated stimuli, then this was evidence for holistic processing. To assess whether our model could account for their observations (some of which were interpreted as indicating parts-based processing), we simulated their experiment with POFA (Ekman & Friesen, 1976) (Figure 7b–c).

Simulation 3: Experiment 3 of Tanaka et al. (2012) Tanaka et al. (2012) looked at the performance of subjects in identifying happy and angry expressions in the lower and upper face halves respectively with two different stimuli types: (1) normal (happy + happy, angry + angry), and (2) angry-happy (happy lower half + angry top half), under two different conditions of alignment: aligned and misaligned (see Fig. 5). We note that their third experiment was based on the partial design which we explained previously has drawbacks. Their reasoning was that equal performance in the aligned

[1] We have chosen to label as normal and angry-happy the stimuli types that they called congruent and incongruent to avoid confusion in this paper.

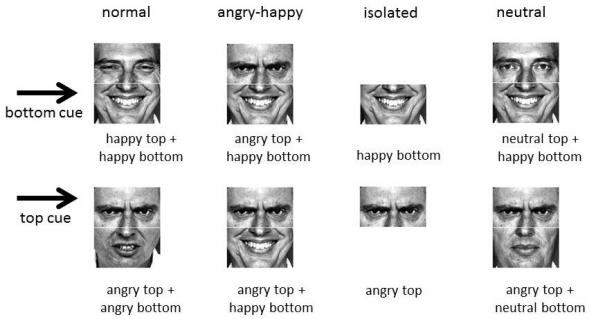

Figure 4: Stimuli types from Exp. 1(Tanaka et al., 2012).

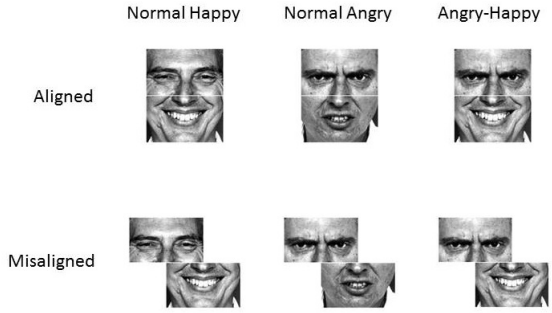

Figure 5: Stimuli types from Exp. 3 (Tanaka et al., 2012).

and misaligned conditions for the normal stimuli would indicate parts-based processing and better performance in the misaligned condition relative to the aligned condition for the angry-happy stimuli would indicate holistic processing. Once again, to assess whether our model could account for their observations (some of which they attributed to parts-based processing), we simulated their experiment with POFA (Ekman & Friesen, 1976).

Results and Discussion

In Simulation 1, the sensitivity, d', for the incongruent trials was less than the sensitivity for the congruent trials (Fig. 6). This observation of the congruency effect confirmed that the model processes facial expressions holistically. It is now left for an experiment in expression recognition with humans based on the CCP to be conducted; we expect that a congruency effect will be observed.

In their first experiment, Tanaka et al. (2012) found that, for happy expressions, the percent accuracy on angry-happy stimuli was reliably less than the accuracy on the other stimuli types. In the case of the model, while there were no statistically significant differences between the four stimuli types, we observed a very similar trend. The reaction time for the angry-happy stimuli was reliably greater than the reaction time for the other three stimuli types in the experiment. In modeling, the difference in reaction time for angry-happy and normal stimuli approached statistical significance, and was reliable for angry-happy and isolated stimuli. Notably, there were no significant differences between the percent accuracy

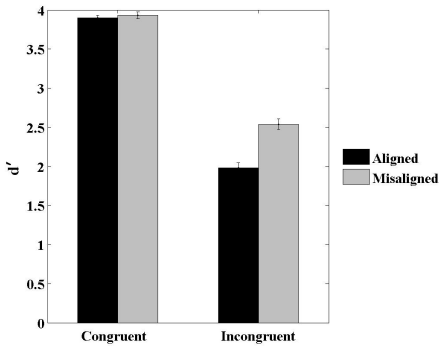

Figure 6: sensitivity (d') in aligned and two misaligned conditions for congruent and incongruent trials of the complete composite paradigm. Error bars indicate standard error of the mean.

or reaction time for the normal, isolated and neutral stimuli in the experiments and in modeling. From their observations, Tanaka et al. (2012) concluded that the recognition of lower half face happy expressions is analytic when there is little or no conflict between the face halves (e.g. normal, isolated and neutral stimuli) but holistic when their is a clash between the face halves (e.g. angry-happy stimuli). However, we made similar observations in the model which possesses a single pathway of holistic processing.

For angry expressions, Tanaka et al. (2012) found that the percent accuracy on isolated stimuli was less than the accuracy on normal and neutral stimuli, and that the accuracy on angry-happy stimuli was less than the accuracy on the other three stimuli types; no other differences were observed for accuracy or reaction time. They concluded from their observations that the recognition of top half face angry expressions was not purely analytic and in fact benefited from the presence of whole-face information. While we did not observe any statistically significant differences between accuracy or reaction time for the four stimuli types in the model, we - once again - observed a similar trend with the model (Fig. 7–8).

In their third experiment, Tanaka et al. (2012) found a lower percent accuracy and greater reaction time for angry-happy stimuli in the aligned condition relative to the misaligned condition, and attributed this to holistic processing. In contrast, for normal stimuli, there was no difference between the accuracy in the aligned and misaligned conditions. They expressed surprise at seeing a shorter reaction time in the misaligned relative to the aligned condition. They attributed the absence of a holistic facililation for aligned normal stimuli to analytic processing. Yet once again, we observed very similar trends in the model, suggesting that holistic processing suffices for explaining the experimental results (Fig. 9–10).

Our modeling work emphasizes the importance of modeling in cognitive science. Without a model, inferences from experimental results can be misleading. While we cannot

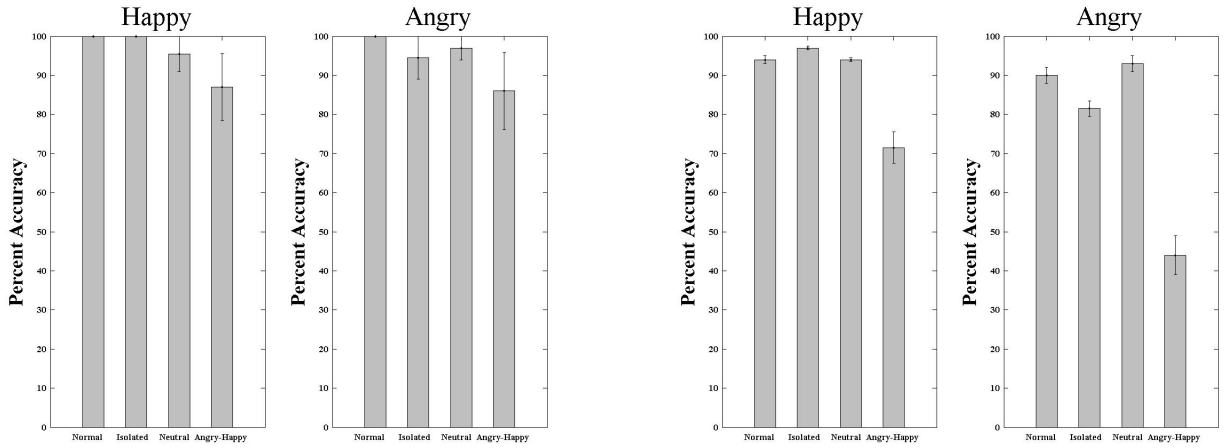

Figure 7: percent accuracy in simulation 2 with our model (a) and experiment 1 in Tanaka et al. (2012) (b).

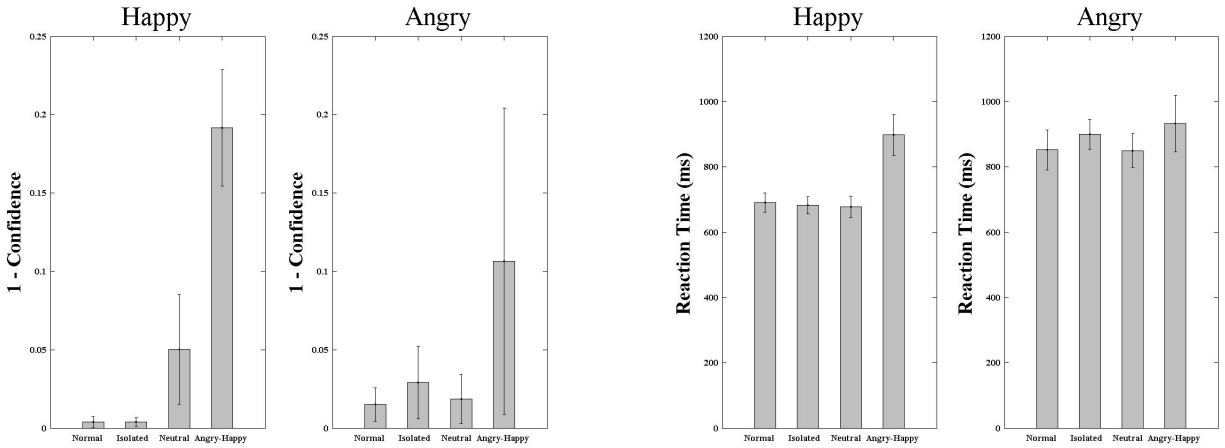

Figure 8: "reaction time" in simulation 2 with our model (a) and reaction time in experiment 1 of (Tanaka et al., 2012) (b).

claim - based on our model - that facial expression processing in humans is purely holistic, we can conclude that Tanaka et al.'s experiments do not show that it has analytic attributes.

Acknowledgments

This work was supported in part by NSF grants SMA-1041755 and IIS-1219252 to G.W. Cottrell. We thank the Perceptual Expertise Network (PEN) and Gary's Unbelievable Research Unit (GURU) for discussion and comments, and Jim Tanaka for permission to use his data.

References

Audley, R. J. (1960). A stochastic model for individual choice behavior. *Psychological Review*, *67*, 1–15.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 929–945.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 527–551.

Cheung, O. S., Richler, J. J., Palmeri, T. J., & Gauthier, I. (2008). Revisiting the role of spatial frequencies in the holistic processing of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1327–1336.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). Empath: a neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, *14*(8), 1158–1173.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-

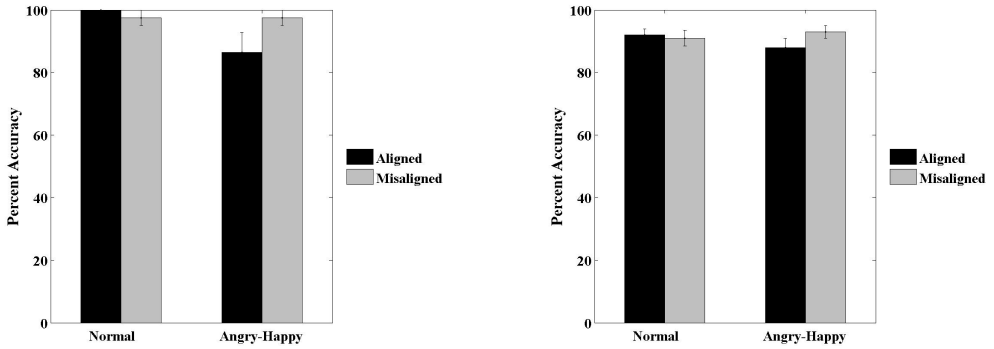

Figure 9: percent accuracy in simulation 3 with our model (a) and experiment 3 of Tanaka et al. (2012) (b).

dimensional visual cortex filters. *Journal of the Optical Society of America A*, *2*, 1160–1169.

Daugman, J. G. (1988). Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *36*(7), 1169–1179.

Dias, S. S. (2007). *Improved 2d gabor filter.* Available from `http://www.mathworks.com/matlabcentral/fileexchange/13776-improved-2d-gabor-filter` (Retrieved May 7, 2011)

Ekman, P., & Friesen, W. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychologists Press.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition*, *103*, 322–330.

Hubel, D. H. (Ed.). (1988). *Eye, brain, and vision*. New York, NY: Scientific American Library.

LeCun, Y., Bottou, L., Orr, G. B., & Mueller, K. Robert. (1998). Efficient backProp. In G. Orr & K. Mueller (Eds.), *Neural networks: tricks of the trade.* New York: Springer.

Movellan, J. R. (2002). *Tutorial on gabor filters* (Tech. Rep.). La Jolla, CA: University of California San Diego.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, *22*(4), 464–471.

Richler, J. J., Tanaka, J. W., Brown, D. D., & Gauthier, I. (2008). Why does selective attention to parts fail in face processing? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 1356–1368.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*, 225–245.

Tanaka, J. W., Kaiser, M. D., Butler, S., & Le Grand, R. (2012). Mixed emotions: holistic and analytic perception of facial expressions. *Cognition & Emotion*, *26*(6), 961–977.

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.

Young, A., Hellawell, D., & Hay, D. (1987). Configural information in face perception. *Perception*, *16*, 747–759.

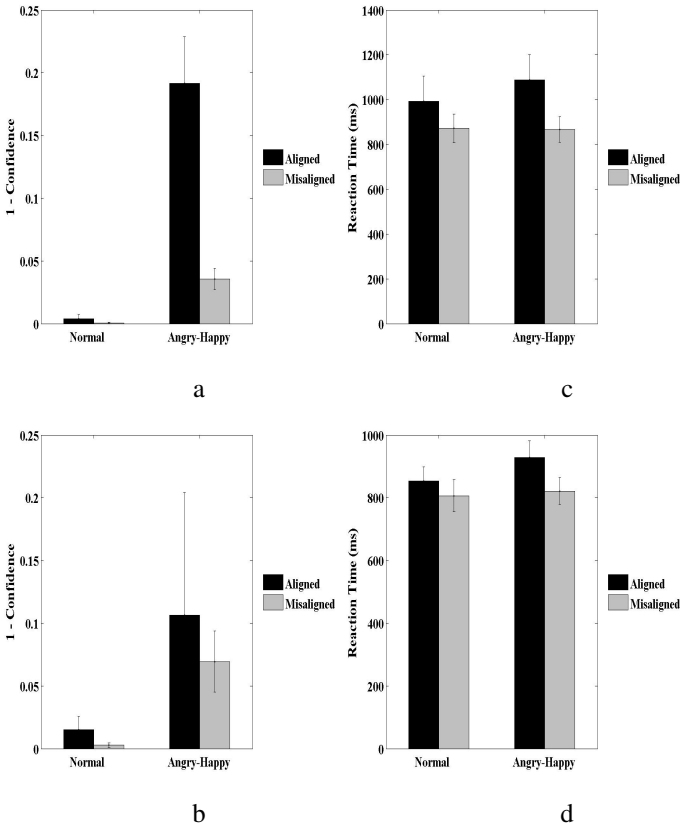

Figure 10: "reaction time" in simulation 3 with our model in identifying happy (a) & angry (b) facial expressions and reaction time in experiment 3 of Tanaka et al. (2012) for happy (c) & angry (d) facial expressions.