Exploring the Facial Expression Perception-Production Link Using Real-Time Automated Facial Expression Recognition

David M. Deriso[1], Josh Susskind[1], Jim Tanaka[2], Piotr Winkielman[3], John Herrington[4], Robert Schultz[4], and Marian Bartlett[1]

[1] Machine Perception Laboratory, University of California, San Diego
dderiso@ucsd.edu, {josh,marni}@mplab.ucsd.edu
[2] Department of Psychology, University of Victoria
[3] Department of Psychology, University of California, San Diego
[4] Center for Autism Research, Children's Hospital of Philadelphia

Abstract. Motor production may play an important role in learning to recognize facial expressions. The present study explores the influence of facial production training on the perception of facial expressions by employing a novel production training intervention built on feedback from automated facial expression recognition. We hypothesized that production training using the automated feedback system would improve an individual's ability to identify dynamic emotional faces. Thirty-four participants were administered a dynamic expression recognition task before and after either interacting with a production training video game called the Emotion Mirror or playing a control video game. Consistent with the prediction that perceptual benefits are tied to expression production, individuals with high engagement in production training improved more than individuals with low engagement or individuals who did not receive production training. These results suggest that the visual-motor associations involved in expression production training are related to perceptual abilities. Additionally, this study demonstrates a novel application of computer vision for real-time facial expression intervention training.

1 Introduction

Facial expression production and perception are crucial components of social functioning. Disruptions in these processes limit an individual's ability to interact with others, which has negative consequences on health and quality of life [1,2]. Neurophysiological disorders such as autism spectrum disorders (ASD) are associated with reduced ability to produce [3] and perceive [4] facial expressions [5]. Recent advances in automated facial expression recognition technology open new possibilities for clinical interventions that target these deficits. The broader goal of this project is to develop an intervention based on facial expression production by leveraging automatic facial expression recognition technology to provide real-time feedback on the participants own expressions.

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part II, LNCS 7584, pp. 270–279, 2012.

The present study employs the automated expression recognition intervention to explore the link between facial expression perception and production. We examine the effect of emotional facial expression production training on the perception of dynamic facial expressions. To the best of our knowledge, this is the first paper to address the question of whether training facial expression production influences perception. Previous computer-based interventions focused on expression recognition [6,7], but not production, and the link between facial production and perception remains unclear. Practice with expression production may not only influence production itself, but may also influence perception due to possible associations between recognition and motor production.

1.1 Exploring the Expression Perception and Production Link

The recruitment of the motor system for perceptual learning is a theme that has become widespread in cognitive science, dating back to the motor theory of speech perception [8]. The theory posits that motor production plays an important role in the development of perceptual recognition capabilities. More recent evidence suggests that perceptual recognition itself involves dynamic interaction through a network of brain areas including the motor system. Neurons in a visual-motor system, called "mirror neurons," have been reported to activate during the production of specific muscle movement and also to visual display (perception) of the same muscle movement performed by another [9]. There is evidence from cognitive neuroscience and psychology that perceptual systems for facial expression recognition may be linked with motor as well as somatosensory feedback systems [10,11]. For example, facial motor interference produced by biting a pen has been shown to impair recognition of certain emotional expressions [12]. Moreover, disrupting sensory feedback from the facial region of somatosensory cortex via TMS impairs discrimination of facial expressions, but not recognition of facial identity [13].

Together, these studies suggest that perception and production may be linked through a common use of the motor system. The present study explores this link by measuring the impact of facial expression production training on the ability to recognize dynamic facial expressions. We hypothesized that training with facial expression production will improve perceptual ability to recognize dynamic facial expressions.

To study the effects of facial expression production, we employed real-time facial expression recognition within a novel intervention game called the emotion mirror. In this paradigm, subjects receive closed-loop sensorimotor feedback, where they produce facial expressions and simultaneously view their expressions mirrored by interactive animations.

2 Methods

To measure the effect of production training on perceptual ability, a pre-test post-test design was employed where perceptual ability was measured using the

Dynamic Expression Recognition Task (DERT) before and after the intervention task (Emotion Mirror) or control task (Sushi Cat).

2.1 Dynamic Expression Recognition Task ("DERT")

Social interactions demand efficient and accurate interpretations of dynamic facial expressions. However, previous studies of face perception have predominantly used static, high intensity facial stimuli, with which participants are often at ceiling recognition performance. To measure the perceptual ability for dynamic facial expressions, DERT was adapted from [14,5].

Participants were shown a human face that gradually morphed over the course of 500 ms from a neutral expression to one of four basic emotions: sad, angry, fearful, and surprised. In order to assess the perceptual limitations for identifying an emotion, the amount of information given for each decision was varied by allowing the morph to reach one of five percentages of completion (intensity): 15 (difficult), 30, 45, 60, and 75 (easy). The lowest intensity is presented first, in a block design. Participants were instructed to identify each expression as quickly and accurately as possible, while response reaction time and accuracy were recorded.

Dynamic stimuli were created by generating morphs of six actors selected from the Radboud database [15] of basic emotional expressions. Each actor's face was morphed from neutral to a high intensity expression using facial action constellations defined in the Facial Action Coding System (FACS) manual [16] as markers. Morphs were rapidly displayed as sequences that lasted a total of one second, beginning with a neutral expression that lasted for 500 msec, followed by a dynamic expression (See Figure 1).

DERT differs from [14,5] in that the dynamics are consistent across expression intensities; it does not slow down the frame rate or rate of expression production for the low intensity expressions, nor does it hold a static facial expression at the end of the dynamic display. Because the low intensity expressions had fewer frames, a one second total display time was achieved by increasing the duration of the static neutral expression before the onset of the low intensity morph (See Figure 1).

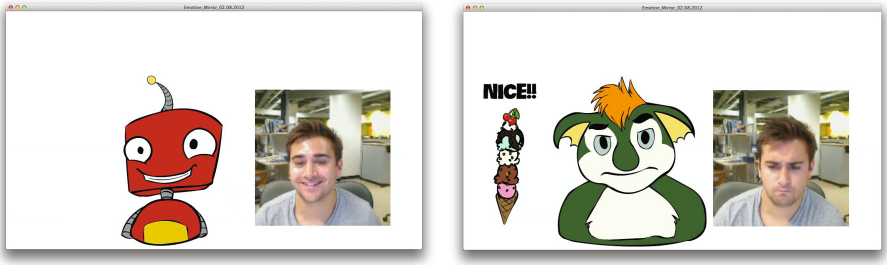
Fig. 1. Dynamic Expression Recognition Task

2.2 The Emotion Mirror

The emotion mirror is an intervention game in which an illustrated character responds to the facial expression of the subject in real-time. There are two conditions that govern the direction of the interaction: "it mimics you" and "you mimic it." In the first condition, the participant is audibly prompted to make one of six expressions. If the correct expression is produced, the character mirrors the participant's expression (ie. "it mimics you"), which is rewarding to the participant. In the second condition, the character displays and holds an expression, and participant is instructed to copy the character (ie. "you mimic it"). As the participant approximates the expression of the character, an ice-cream cone grows by increasing in units of scoops and toppings. This playful interaction is designed to appeal to children and adults.

The game includes a selection of engaging animated characters that range in visual complexity. It is well known that children with ASD tend to be more comfortable with visually simple displays, and are more comfortable with robots and animals than with human faces. Accordingly, the avatars in the intervention game range from animals, to outer-space creatures, to a more realistic human avatar for which children can choose gender, skin color, and hair color to match their own (See Figure 2). Each condition cycled through three characters (dog, girl, boy) for six emotions each (two conditions with three repetitions of six emotions). The task knits expressive production and perception as it is the participant's own face that drives the expressions shown on the avatar, and provides feedback and practice in facial expression mirroring behaviors.

The facial expression recognition engine behind the Emotion Mirror is the Computer Expression Recognition Toolbox (CERT).



(a) It Mimics You (b) You Mimic It

Fig. 2. The Emotion Mirror

2.3 The Computer Expression Recognition Toolbox ("CERT")

CERT is a computer vision system for measuring facial expressions in real time [17,18] (See Figure 3). CERT automatically codes the intensity of six expressions of basic emotion, three dimensions of head pose (yaw, pitch, and roll, as well as

the 20 facial actions from the FACS [8] most related to emotion. FACS is a system for objectively scoring facial expressions in terms of elemental movements, called action units (AUs), which roughly correspond with individual facial muscle movements. FACS provides a comprehensive description of facial expressions, providing greater specificity and diversity than emotion categories.

The technical approach to CERT is an appearance-based discriminative approach. Such approaches have proven highly robust and fast for face detection and tracking (e.g. [19], and take advantage of the rich texture information in facial expression images. This class of approaches achieves a high level of robustness through the use of large data sets for machine learning. This speed and robustness has enabled real-time analysis of spontaneous expressions. Evaluation of CERT on benchmark datasets shows state-of-the-art performance for both recognition of basic emotions and recognition of facial actions. CERT also provides information about expression intensity. See [17] for more information on system design and benchmark performance measures. CERT operates in near real-time at approximately 12-15 frames per second, and is available for academic use.

A softmax competition between CERT's six expressions of basic emotion and neutral mediates the matching signal between the prompted emotion and that of the participant. This matching signal drives the character's expression intensity or the height of the ice cream in proportion to the participant's expression intensity.

2.4 Participants

Participants were thirty-four healthy college students (M = 14, F= 19, M(age) = 20.5) who volunteered for course credit. All volunteers were instructed to bring corrective eyewear if needed. No subjects were excluded, although data from four intervention subjects sets were unusable due to recording errors. The study was approved by the Human Research Protections Program at the University of California, San Diego.

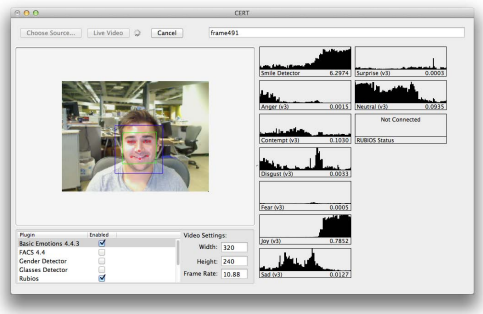


Fig. 3. The Computer Expression Recognition Toolbox ("CERT")

2.5 Experimental Design

Each participant was randomly assigned into one of two groups: intervention (N=17) and control (N=17). The intervention group received facial expression production practice, and the control group played a spatial game that does not involve expression perception or production. All participants were assessed for recognition of dynamic facial expressions before and after intervention using the Dynamic Expression Recognition (DERT) task. The intervention group performed facial expression production exercises for 20 minutes, starting first with the Emotion Mirror "it mimics you," followed by "you mimic it," and finally Face-Face-Revolution [7]. The control group played a spatial computer game that does not involve expression perception or production called "Sushi Cat" [908] for 20 minutes. The object of the game is to drop a cat from an optimal position such that as the cat falls and bounces off obstacles, it can eat the maximum amount of sushi before landing. The control task does not prompt the user to produce facial expressions.

2.6 Measures

The primary dependent measures for DERT were accuracy, and reaction time. Pre-test scores were used as a covariate to compare post-test scores across intervention and control groups. Confusion matrices were also analyzed by comparing their similarity to the identity matrix using the Backus-Gilbert spread (BGS) (Equation. (1)) [20]. The BGS measures the Euclidean distance between two vectorized matrices, normalized by the number of off-diagonal elements.

$$1 - \left[\sum_{ij} \left(X_{ij} - I_{ij} \right)^2 / \sum_{mn} \left(1 - I_{mn} \right)^2 \right] \tag{1}$$

In order to determine the extent to which participants made an effort to undergo production training with the emotion mirror, a measure of engagement, or compliance, was assessed using the mean accuracy of expression production. To calculate this, a true positive was awarded for a given trial when the peak emotion measured by CERT (integrated across the duration of the trial) matched the target emotion prompted by the Emotion Mirror.

Speed of facial expression production during the emotion mirror exercise was also assessed as the time from the stimulus onset to the peak CERT value for the target emotion (time-to-peak).

3 Results

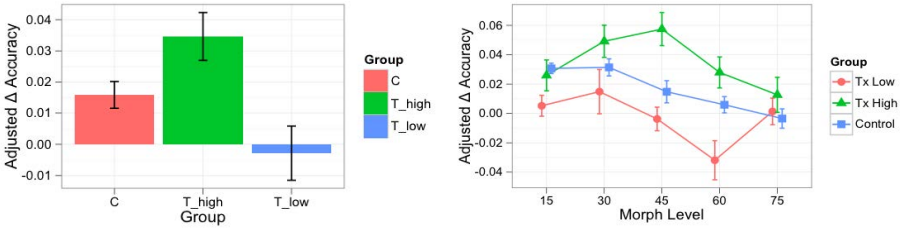
We tested the hypothesis that engagement with the emotion mirror intervention (EM) would lead to improvements in expression recognition as measured by the DERT. As predicted, there was a trend for a positive correlation between engagement with the Emotion Mirror and gain scores on the dynamic expression recognition task ($r = .517, n = 13, t(1, 11) = 2.01, p < 0.07$) as measured by a Pearson

product moment correlation. The correlation was stronger for the Emotion Mirror's "you mimic it" condition ($r = .612, n = 13, t(1, 11) = 2.57, p < 0.026$). Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity, and homoscedasticity.

To investigate whether the perceptual improvement was related to intervention engagement or chance, the intervention group was split by the median level of engagement into two groups: high engagement and low engagement (abbreviated as "T_high" and "T_low"). The pre and post-test performance of these new groups were then compared against a control group, who were given the same tests.

Since our main dependent measure of accuracy has a fixed ceiling at 1.00, we employed the ANCOVA method to improve the sensitivity of our between-group comparisons [21]. A two-way mixed analysis of covariance was conducted to assess the effectiveness of the emotion mirror intervention in improving accuracy on the dynamic expression recognition task. The within-subject independent variable was the amount of facial morph (15, 30, 45, 60, 75) and the between-subject independent variable was group assignment (control, intervention with high EM accuracy, intervention with low EM accuracy). The dependent variable was pos-test DERT accuracy scores, while DERT pre-test scores were used as a covariate to control for individual differences. Checks were conducted to ensure that there were no violations of the assumptions of normality, linearity, homogeneity of variance, homogeneity of regression slopes, and reliable measurement of the covariate. After adjusting for pre-test DERT accuracy scores, there was a significant main effect of treatment group ($F(1, 2) = 3.361, p < 0.006$). There was also a significant main effect of morph level ($F(1, 4) = 13.812, p < .00001$), and an interaction of treatment group by morph level ($F(1, 8) = 3.357, p < 0.002$) (See Figure 4).

To assess whether the intervention group improved more than the control group, we compared the accuracy improvement (gain score) for each group collapsed across levels of morph. Post-hoc Welch independent two sample t-tests were performed on the adjusted post-test score. These tests revealed that the intervention group with high EM accuracy had significantly greater improvements than did the control group ($t(55.8) = -3.03, p < 0.00373$, two-sided, mean difference = 0.02, 95% CI: -0.0312 to -0.00635). These tests also revealed that the intervention group with high EM accuracy had significantly greater improvements than did the intervention group with low EM accuracy ($t(62.6) = 4.89, p < .00001$, two-sided, mean difference = -0.038, 95% CI: 0.02 to 0.05), and the control group had significantly greater improvements than did the intervention group with low EM accuracy ($t(48) = 3.02, p < 0.004$, two-sided, mean difference = -0.019, 95% CI: 0.006 to 0.031). These results suggest that changes in perceptual accuracy from pre-test to post-test is strongly influenced by the accuracy during the emotion mirror task. These findings support the influence of expression practice and visual-motor associations, in expression recognition.



(a) Across Morph (b) By Morph

Fig. 4. Perceptual Improvement (Post - Pre)

The relationship between the latency of production (time-to-peak) and improvement in DERT reaction time was assessed with a Pearson product moment correlation. The quicker an individual was to maximally produce the correct expression was significantly related to quicker reaction times ($r = -0.558, n = 13, t(1, 11) = -2.23, p < 0.048$). In addition, confusion matrices were quantified by measuring the similarity to the identity matrix using the Baccus-Gilbert Spread (BGS). Improvement in the BGS during the emotion mirror was significantly correlated in improvement in the BGS for recognition as measured by DERT ($r = .558, n = 13, t(1, 11) = 2.23, p < 0.048$).

4 Discussion

Consistent with the hypothesis that practice with production leads to improvements in perception, we found that greater engagement with the emotion mirror intervention improves perceptual accuracy for dynamic faces across emotions and varying levels of difficulty. The perceptual benefit was tied to expression production performance during motor training. These findings are consistent with our initial theory that the visual-motor associations gained in production training influence perceptual abilities.

In further support of the role of facial motor practice in perception, we found that the increase in speed of expression production during the Emotion Mirror was correlated with recognition performance improvement in DERT. This result demonstrates a relationship between learning in facial motor production and learning in facial expression perception that cannot be accounted for by practice recognizing facial expressions alone, and supports the hypothesis that the motor production itself contributed to perceptual improvement.

Furthermore, improvements in the confusion matrix (measured by the BGS) for facial expression production during the Emotion mirror were significantly correlated in improvement in the BGS for recognition as measured by DERT, providing further evidence that improvements in production are related to improvements in perception.

Though no significant differences were observed between engagement with the two production conditions, the "it mimics you" condition showed a stronger correlation with perceptual improvement in DERT than did "you mimic it." Though the differences between these two tasks were not explicitly studied, this correlation suggests that there may be a difference in the efficacy between the two methods of training. Further studies are needed to tease apart whether this correlational discrepancy was due to differences in the way each condition trained the participant, or if one condition had a stronger coupling between the subject and the emotion mirror

This study is not without limitations. Facial expression production training can be a tiring process, and the current methods were unable to adjust for how tired the participants were when they underwent the second, also tiring, perceptual assessment. Not all participants may have felt comfortable producing facial expressions, and some may have felt embarrassed more so than would children who may be more engaged with the illustrated characters and ice cream. CERT does not necessary work equally well for all participants, and some may have had a harder time engaging that others.

The present study is a product of a multi-disciplinary research effort to develop and evaluate a computer-aided intervention system to enhance the facial expression skills of children with ASD that merges the expertise of researchers in computer vision, face perception, autism, and social neuroscience. Recent advances in computer vision open new avenues for computer assisted intervention programs that target critical skills for social interaction, including the timing, morphology and dynamics of facial expressions. Such systems enable investigations into the learning of facial expression production that were previously not possible. The Emotion Mirror presents pioneering work to develop an engaging intervention system for facial expression processing based on automatic facial expression recognition technology. Such technology contributes not only to potential treatments, but also to the study of learning and plasticity in perception and production systems, and to understanding the cognitive neuroscience of emotion.

Acknowledgment. Support for this work was provided by NIH grant NIMH-RC1 MH088633 and NSF grant SBE-0542013. Any opinions, findings, conclusions or recommendations expressed in this material do not necessarily reflect the views of the National Science Foundation.

References

1. Eisenberger, N., Cole, S.: Social neuroscience and health: neurophysiological mechanisms linking social ties with physical health. Nature Neuroscience (2012)
2. House, J., Landis, K., Umberson, D.: Social relationships and health. Science 241, 540–545 (1988)
3. McIntosh, D., Reichmann-Decker, A., Winkielman, P., Wilbarger, J.: When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. Developmental Science 9, 295–302 (2006)

4. Adolphs, R., Sears, L., Piven, J.: Abnormal processing of social information from faces in autism. Journal of Cognitive Neuroscience 13, 232–240 (2001)
5. Rump, K., Giovannelli, J., Minshew, N., Strauss, M.: The development of emotion recognition in individuals with autism. Child Development 80, 1434–1447 (2009)
6. Madsen, M., El Kaliouby, R., Goodwin, M., Picard, R.: Technology for just-in-time in-situ learning of facial affect for persons diagnosed with an autism spectrum disorder. In: Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 19–26. ACM (2008)
7. Tanaka, J., Wolf, J., Klaiman, C., Koenig, K., Cockburn, J., Herlihy, L., Brown, C., Stahl, S., Kaiser, M., Schultz, R.: Using computerized games to teach face recognition skills to children with autism spectrum disorder: the lets face it! program. Journal of Child Psychology and Psychiatry 51, 944–952 (2010)
8. Liberman, A., Mattingly, I.: The motor theory of speech perception revised. Cognition 21, 1–36 (1985)
9. Rizzolatti, G., Craighero, L.: The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192 (2004)
10. Heberlein, A., Adolphs, R.: Neurobiology of emotion recognition. Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior 31 (2007)
11. Niedenthal, P., Mermillod, M., Maringer, M., Hess, U.: The simulation of smiles (sims) model: Embodied simulation and the meaning of facial expression. Behavioral and Brain Sciences 33, 417–433 (2010)
12. Oberman, L., Winkielman, P., Ramachandran, V.: Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. Social Neuroscience 2, 167–178 (2007)
13. Pitcher, D., Garrido, L., Walsh, V., Duchaine, B.: Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. The Journal of Neuroscience 28, 8929–8933 (2008)
14. Montagne, B., Kessels, R., De Haan, E., Perrett, D.: The emotion recognition task: A paradigm to measure the perception of facial emotional expressions at different intensities. Perceptual and Motor Skills 104, 589–598 (2007)
15. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D., Hawk, S., Van Knippenberg, A.: Presentation and validation of the radboud faces database. Cognition and Emotion 24, 1377–1388 (2010)
16. Ekman, P., Friesen, W.: Facial action coding system: A technique for the measurement of facial movement (1978)
17. Galantucci, B., Fowler, C., Turvey, M.: The motor theory of speech perception reviewed. Psychonomic Bulletin & Review 13, 361–377 (2006)
18. Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., Bartlett, M.: The computer expression recognition toolbox (cert). In: 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), pp. 298–305. IEEE (2011)
19. Pelc, K., Kornreich, C., Foisy, M., Dan, B.: Recognition of emotional facial expressions in attention-deficit hyperactivity disorder. Pediatric Neurology 35, 93–97 (2006)
20. Schyns, P., Petro, L., Smith, M.: Transmission of facial expressions of emotion coevolved with their efficient decoding in the brain: behavioral and brain evidence. PLoS One 4, e5625 (2009)
21. Cribbie, R., Jamieson, J.: Decreases in posttest variance and the measurement of change. Methods of Psychological Research Online 9, 37–55 (2004)