Title
Facial Motor Information is Sufficient for Identity Recognition

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Vitale, Jonathan
Johnston, Benjamin
Williams, Mary-Anne

Publication Date
2017

Peer reviewed

Facial Motor Information is Sufficient for Identity Recognition

Jonathan Vitale, **Benjamin Johnston**, & **Mary-Anne Williams**
{**jonathan.vitale, benjamin.johnston, mary-anne.williams**}**@uts.edu.au**
University of Technology Sydney – Centre For Artificial Intelligence
Innovation and Enterprise Research Lab (The Magic Lab) - 15 Broadway, Ultimo NSW 2007 - Australia

Abstract

The face is a central communication channel providing information about the identities of our interaction partners and their potential mental states expressed by motor configurations. Although it is well known that infants ability to recognise people follows a developmental process, it is still an open question how face identity recognition skills can develop and, in particular, how facial expression and identity processing potentially interact during this developmental process. We propose that by acquiring information of the facial motor configuration observed from face stimuli encountered throughout development would be sufficient to develop a face-space representation. This representation encodes the observed face stimuli as points of a multidimensional psychological space able to assist facial identity and expression recognition. We validate our hypothesis through computational simulations and we suggest potential implications of this understanding with respect to the available findings in face processing.

Keywords: face perception; face processing; face-space; face identity processing; face expression processing; mirroring

Introduction

Face processing capabilities are of paramount importance for the development of social skills (Grossmann, 2015).

Developmental studies suggest that newborns can match observed facial motor configurations via overt imitative behaviour (Meltzoff & Moore, 1983, 1992) or covert inner simulation mechanisms (Simpson, Murray, Paukner, & Ferrari, 2014; Gallese & Caruana, 2016), even well before the development of early cognitive capabilities (but see Oostenbroek et al., 2016 and Simpson et al., 2016 for a recent discussion on the topic). Hence, it has been suggested that facial expression recognition may be mediated by early neural mechanisms mapping sensory information of the observed facial configuration into a proprioceptive motor format (Gallese & Caruana, 2016; Iacoboni, 2009) and therefore assisting imitatory mechanisms (Simpson et al., 2014).

On the contrary, face identity processing capabilities follow a developmental process (Grossmann & Vaish, 2009). Currently, facial identity processing development is not yet well understood. For example, we do not know yet where in the face processing hierarchy representations of invariant (*i.e.* identity features of the face) and dynamic (*i.e.* motor features of the face) features interact (Simion & Di Giorgio, 2015).

According to the *'face-space'* framework (Valentine, 1991; Valentine, Lewis, & Hills, 2015), facial representations are encoded in a multidimensional psychological space. The dimensions of this space are assumed to encode properties of the facial signals that better discriminate one face from another. The distance between two representations underlies their dissimilarity from a psychological perspective. This framework was initially designed to only account for coding *identity-related* features, such as sex, distinctiveness, age and attractiveness (Valentine, 1991). Nevertheless, dynamic aspects of faces, such as facial expressions, were neglected. Recently, we developed a computational tool building on top of the face-space framework (Vitale, Williams, & Jonhston, 2016) and able to exhibit interesting features in agreement with modern understanding in face processing studies. In particular, we demonstrated that this novel face-space can represent both invariant and dynamic features of face stimuli under a shared representation facilitating the recognition of both facial expression and identity exhibited by novel face stimuli (Vitale et al., 2016).

In this paper we offer a new understanding of this face-space, suggesting that facial identity processing capabilities can plausibly develop by interpreting the motor configuration of observed face stimuli.

In particular, from a functional level of analysis, we aim to demonstrate that assuming the existence of an early or innate system $\mathcal{M}otor(x_i) \Rightarrow \mathcal{E}(x_i)$ able to map perceptual information of the observed face stimulus x_i onto a motor interpretation of the exhibited facial expression $\mathcal{E}(x_i)$, it is possible to develop another system $\mathcal{C}ognitive(X_{new}) \Rightarrow \{\mathcal{E}(X_{new}), \mathcal{I}(X_{new})\}$ assisting the discrimination of facial expressions $\mathcal{E}(X_{new})$ and identities $\mathcal{I}(X_{new})$ exhibited by newly encountered face stimuli X_{new}. Therefore, this paper aims to provide computational evidence supporting the following hypothesis:

Hypothesis: It is possible to generalise the face-space framework to realise a twofold multidimensional space structure able to facilitate facial expression and identity processing capabilities by only interpreting the motor configuration exhibited by the face stimuli encountered during the developmental process.

This work is a significant contribution able to provide a plausible explanation unifying traditional and modern findings in face processing studies, as we will discuss in the remainder of this paper.

Previous Findings

Recently, we provided a novel understanding of the face-space framework (Vitale et al., 2016). The face-space framework is a widely used tool in face perception and processing research able to explain many of the phenomena underlying facial identity discrimination in both human experimental settings (Lee, Byatt, & Rhodes, 2000; Rhodes, Jaquet, et

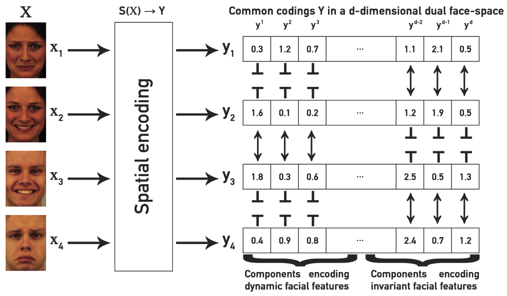


Figure 1: The dual face-space presents a twofold structure: on one side it allows observations with similar motor configurations to lie within close spatial locations (↕), whereas at the same time "repulsing" observations of similar identities away (⊣⊢); on the other side, it happens exactly the viceversa. This facilitates respectively facial expression and identity recognition, under common multidimensional codings.

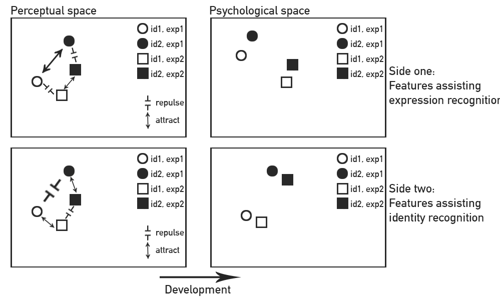


Figure 2: An example of face-space development resulting by applying the mapping function in Equation 2. Face samples belonging to the same identity are on average perceptually closer to each other, thus being a bias for the classification of facial expressions.

al., 2011) and computational simulations (A. J. Calder, Burton, Miller, Young, & Akamatsu, 2001). This framework is so important in face studies that it is *"virtually impossible to explain the interactions between the computational and cognitive approaches to understanding face recognition without reference to this model. It serves as the glue that binds the theoretical and computational aspects of the problem together"* (A. Calder, 2011, page 17).

According to Valentine's face-space, faces are points of a multidimensional space based on their perceived properties. This structure can plausibly account for coding *identity-related* features. Unfortunately, dynamic aspects of the face, such as its motor configuration, were neglected in the traditional face-space account. This is a significant limitation, preventing the analysis of the interactions happening between facial expression and facial identity processing.

Therefore, to fill this gap, we introduced a novel hypothesis: the *duality hypothesis*. This hypothesis suggests that the face-space can plausibly exhibit a twofold structure integrating both dynamic and invariant features of the face into shared codings, although preserving some separation among them to facilitate both facial expression and identity recognition (see Figure 1 for a visual example). We named this understanding with *dual face-space* and we validated the hypothesis, from a computational perspective, through a mathematical presentation and quantitative results.

The Dual Face-Space

Given a set of face stimuli shaped as column vectors of a matrix X, these stimuli have dimension $\mathscr{D}$ equal to the total number of pixels representing each face stimulus. By submitting the matrix X to a Principal Component Analysis (PCA) (Turk & Pentland, 1991) it is possible to obtain a mapping matrix V_{pca} able to map the $\mathscr{D}$-dimensional face stimuli X into compressed d-dimensional representations $\bar{X}$. This process preserves most of the information carried by the face stimuli, but it compresses them in representations having dimension $d \ll \mathscr{D}$ and it ensures desirable properties in subsequent stages of the model (*e.g.* positive definiteness, see Vitale et al., 2016):

$$\bar{X} = V_{pca}^{\top} X \tag{1}$$

It is important to note that in this paper we do not aim to test the classification performance of the proposed model against other computational models of face recognition, but rather the plausibility of the proposed hypothesis in providing a new understanding of the mechanisms potentially underlying human face processing skills. Therefore, in our studies we used the pixels intensities of static images as input to our models to provide a simplified linear understanding of our theory and related argument. Importantly, the input $\bar{X}$ can be any vector of features extracted by the given face stimuli and able to encode perceptual information of the observed stimuli. Therefore, a viable non-linear alternative of our model can be obtained by pre-processing the input face stimuli X by using an unsupervised deep neural network model trained to preserve invariant and dynamic features of the face in a more compressed and smart representation (Le et al., 2013), instead of the proposed linear PCA. Finally, temporal dynamics can be included by pre-processing a set of consecutive stimuli instead of static images, or by using other techniques improving temporal coherence in the resulting pre-processed representation (Mobahi, Collobert, & Weston, 2009). These computational pre-processing stages resemble early processing of human visual cortex and are therefore suitable examples for potential future extensions of our theory and related model.

In our previous work (Vitale et al., 2016), we showed that it is possible to implement the dual face-space by solving the following objective function:

$$V^{\star} = \underset{V \in \mathbf{R}^{d \times d}}{\arg\min} \frac{Tr(V^{\top}\bar{X}(I_N - W^{\mathscr{E}})\bar{X}^{\top}V)}{Tr(V^{\top}\bar{X}(I_N - W^{\mathscr{I}})\bar{X}^{\top}V)} \tag{2}$$

where $W^{\mathscr{E}}$ and $W^{\mathscr{I}}$ are two weight matrices setting desired topological constraints on the face-space via the resulting objective mapping matrix $V^{\star}$. It is possible to obtain the weight matrix $W^{\mathscr{E}}$ by knowing the facial expressions exhibited by the training samples and, when this matrix is used in Equation 2, it encourages pairs of samples associated with the same facial expression to be in nearby locations in the resulting face-space:

$$W_{ij}^{\mathscr{E}} = \begin{cases} \frac{1}{n_{\mathscr{E}_i}}, & \text{if } \mathscr{E}(x_i) = \mathscr{E}(x_j) \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

In Equation 3, $n_{\mathscr{E}_i}$ is the number of samples in X belonging to the facial expression class $\mathscr{E}(x_i)$ of the face stimulus x_i in the column i of matrix X.

It is possible to realise the weight matrix $W^{\mathscr{I}}$ by knowing the identities exhibited by the training samples and, when this matrix is used in Equation 2, it promotes repulsive forces between pairs of samples belonging to the same identity, thus reducing misclassification of facial expressions due to the *identity bias* (Sariyanidi, Gunes, & Cavallaro, 2015):

$$W_{ij}^{\mathscr{I}} = \begin{cases} \frac{1}{n_{\mathscr{I}_i}}, & \text{if } \mathscr{I}(x_i) = \mathscr{I}(x_j) \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

In Equation 4, $n_{\mathscr{I}_i}$ is the number of samples in X belonging to the identity class $\mathscr{I}(x_i)$ of the face stimulus x_i in the column i of matrix X. Figure 1 and Figure 2 show examples of the rationale behind the constraints set by the suggested weight matrices in Equation 2.

Finally, given a generic matrix M and the following permutation function:

$$\tilde{M} = \sigma(M) = \begin{pmatrix} m^1 & m^2 & m^3 & \dots & m^d \\ m^d & m^{d-1} & m^{d-2} & \dots & m^1 \end{pmatrix} \tag{5}$$

permutating each column vector m^i with $i \in [1, \dots, d]$ of the matrix M in the inverse order[1] we demonstrated that Equation 2 is sufficient to provide multidimensional representations able to facilitate both facial identity and expression recognition (Vitale et al., 2016).

In fact, given $V^\star$ as the optimal solution of the objective function in Equation 2, we demonstrated that the mapping matrix $\tilde{V}^\star = \sigma(V^\star)$ is the optimal solution of another objective function promoting facial identity discrimination obtained by inverting Equation 2. The mapping matrix $\tilde{V}^\star$ is dual to the mapping matrix $V^\star$, since it shares the same components (*i.e.* column vectors) of $V^\star$ but sorted in the opposite order. Therefore, the objective function in Equation 2 realises common codings able to facilitate on one hand facial expression classification ($V^\star$), and on the other hand facial identity discrimination ($\tilde{V}^\star$).

The Δ Face-Space

To validate our hypothesis, we suggest to approximate the weight matrix $W^{\mathscr{I}}$ with another weight matrix W^{Δ} implemented without necessarily knowing the identity classes of the training face stimuli. In this way the weight matrix $W^{\mathscr{I}}$ in Equation 2 can be replaced by the matrix W^{Δ}, thus realising the following objective function:

$$V^{\Delta\star} = \underset{V \in \mathbf{R}^{d \times d}}{\arg\min} \frac{Tr(V^\top \bar{X}(I_N - W^{\mathscr{E}})\bar{X}^\top V)}{Tr(V^\top \bar{X}(I_N - W^{\Delta})\bar{X}^\top V)} \tag{6}$$

The optimal solution of the objective function in Equation 6 is the mapping matrix $V^{\Delta\star}$. Thus, given a mapping matrix V_{pca} gathered by submitting the training data X to a PCA, as previously described, it is possible to obtain the final mapping matrix $V^{\Delta}_{overall}$ realising the Δ face-space as following:

$$V^{\Delta}_{overall} = V_{pca} V^{\Delta\star} \tag{7}$$

The mapping matrix $V^{\Delta}_{overall}$ is able to realise face-space representations facilitating facial expression recognition, whereas the mapping matrix $\tilde{V}^{\Delta}_{overall} = \sigma(V^{\Delta}_{overall})$, having the same component of $V^{\Delta}_{overall}$ but permutated in the inverse order, realises representations able to facilitate facial identity discrimination, although without the need of knowing the identities exhibited by the training samples, as suggested by our hypothesis.

Defining the New Weight Matrix

The purpose of the weight matrix $W^{\mathscr{I}}$ in Equation 2 is to avoid that two face stimuli sharing the same identity, but exhibiting different facial expressions, would get projected to nearby locations of the face-space promoting their misclassification in the same facial expression class (see Figure 2). This misclassification can easily happen since face stimuli of the same identity share most of their perceptual features, and, on average, they are close-by in the perceptual space (Sariyanidi et al., 2015; Turk & Pentland, 1991). This property exhibited by face stimuli can be used to our advantage to realise the desired weight matrix W^{Δ}.

For each of the N training face stimuli x_i, shaped as column vectors $i \in [1, \dots, N]$ of the matrix X, we denote with Δ_{x_i} the set containing the perceptual distances $\delta(x_i, x_j)$ between the face stimuli x_i and the face stimulus $x_j \in X$ with $i \neq j$ and exhibiting a different facial expression from the one exhibited by x_i:

$$\Delta_{x_i} = \{\delta(x_i, x_j) \mid x_j \in X \wedge x_i \neq x_j \wedge \mathscr{E}(x_j) \neq \mathscr{E}(x_i)\} \tag{8}$$

Since face stimuli of the same identity are perceptually close, their respective distances would be, at least on average, well below their distances from face stimuli with different identities. Then, given the mean $\mu_{\Delta_{x_i}}$ and standard deviation $\sigma_{\Delta_{x_i}}$ of the distances included in the set Δ_{x_i} it is possible to compute the set $\mathscr{I}_i^{\approx}$ described as follow:

$$\mathscr{I}_i^{\approx} = \{x_j \mid \delta(x_i, x_j) < \mu_{\Delta_{x_i}} - \beta\sigma_{\Delta_{x_i}}\} \tag{9}$$

where β is a parameter suggesting how many standard deviations below the mean distance would be set the maximum threshold. In this work, β was set equal to 2.5 after empirical tests with face stimuli gathered from different datasets

[1] In this paper we will use the notation $\tilde{M}$ to denote a matrix having the same column vectors of another matrix M, but sorted in an inverse order.

available in face recognition literature. The resulting set $\mathscr{I}_i^{\approx}$ includes most of the training samples sharing the same identity of the sample x_i.

Therefore, the weight matrix W^{Δ} can be realised as follow:

$$W_{ij}^{\Delta} = \begin{cases} \frac{1}{n_{\cup_{ij}}}, & \text{if } x_j \in \mathscr{I}_i^{\approx} \vee x_i \in \mathscr{I}_j^{\approx} \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

where $n_{\cup_{ij}}$ is the number of unique samples in the set $\mathscr{I}_i^{\approx} \cup \mathscr{I}_j^{\approx}$. The realised weight matrix W^{Δ} is clearly symmetric and the associated Laplacian behaves as a block centring matrix, thus promoting a *norm-based space* (for in-depth details and mathematical proof refer to Vitale et al. (2016)). The objective function in Equation 6 can be solved through the iterative algorithm proposed by Ngo, Bellalij, and Saad (2012), similarly to our previous contribution (Vitale et al., 2016).

Experiments

In this paper, we will evaluate the proposed model using the Karolinska Directed Emotional Faces (KDEF) dataset (Lundqvist, Flykt, & Öhman, 1998), similarly to our previous contribution. The dataset contains static images of 70 subjects—35 female and 35 male—exhibiting seven different prototypical facial expressions of basic emotions (anger, disgust, fear, happiness, neutral, sadness and surprise). The pictures are taken in various face orientations and in two different sessions (A and B).

We used the frontal pictures taken in session A. We extracted the facial region from the images and reduced their resolution to 80×80 pixels. Eyes and mouth were at approximately the same position. Illumination variations were reduced by applying a simple equalisation process to the images (using the histeq function available in Matlab software).

We first pre-processed the data by submitting the pixels of the images in input to a PCA as explained previously. We retained the components able to explain 95% of the variance of the original data resulting in 200 components.

Procedure

The present experiment tests the ability of the new Δ face-space, implemented without knowing the identity labels of the training stimuli, to support subsequent processes of identity and facial expression recognition.

In both the two conditions (*i.e.* facial expression and identity recognition) we used repeated random iterations of the dataset's samples (in this work 35 iterations for both the tasks). In each iteration 25 subjects were randomly selected as the test set among the 70 possible subjects to simulate unfamiliar identities. For each of the 25 selected subjects were randomly chosen 2 facial expressions as probes for the identity recognition task, and the remaining 5 facial expressions as test samples, leading to a total of 125 test samples for each iteration. The images of the other 45 subjects, together with the 50 selected probes, were used as the training set of the current iteration, leading to 365 training samples for each iteration.

With each training data we estimated the mapping matrix $V_{overall}^{\Delta}$ of the Δ face-space proposed in this chapter as per Equations 6 and 7. Then, each test sample was mapped onto the Δ face-space, thus obtaining the encodings $Y^{\Delta\mathscr{E}} = V_{overall}^{\Delta\top} X$ and $Y^{\Delta\mathscr{I}} = \tilde{Y}^{\Delta\mathscr{E}} = \tilde{V}_{overall}^{\Delta\top} X$, respectively used during the expression and identity recognition tasks for the Δ face-space condition.

For each iteration, we compared the performance of the Δ face-space against a baseline approach. The baseline approach used all the pixels of the face stimuli to match similar facial expressions or identities. This is a fair methodology considering we pre-processed raw pixels data with a simple PCA. In our previous contribution (Vitale et al., 2016) we showed that the baseline and PCA performance are not differing. Thus, we used this approach as our baseline to demonstrate that matching the expressions and identities of the considered dataset samples in the perceptual space was not a trivial task and that our psychological face-space can indeed facilitate facial expression and identity recognition.

The classification was performed using the nearest neighbour algorithm. For each sample, x_i, used by the baseline approach, and y_i^{Δ}, used by the face-space model, we computed the Euclidean distances from the centroids of each class in the corresponding space, and we selected the class associated with the centroid closer to the sample.

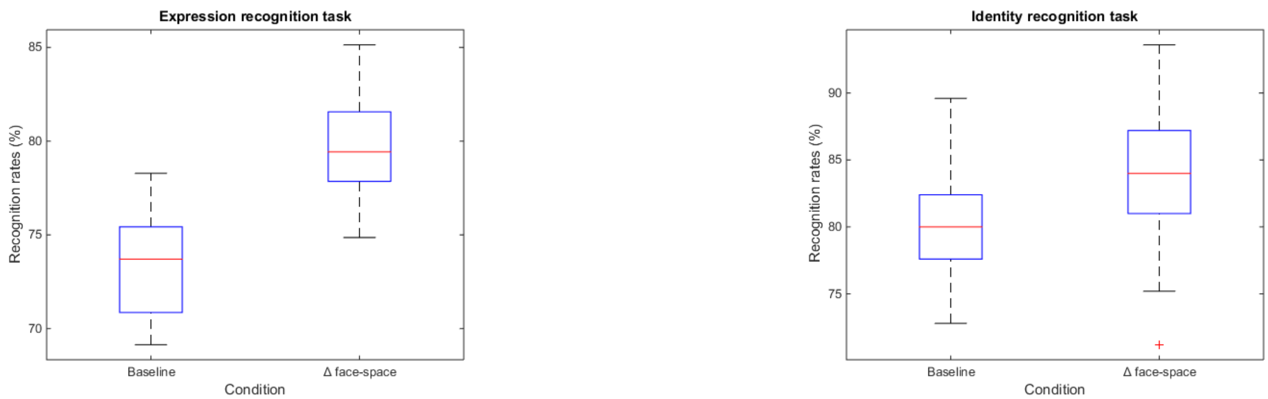
For each test sample during each iteration, the baseline approach provided a single prediction. Instead, our face-space model can use the first $k = [1, \ldots, d]$ components of the mapping matrix $V_{overall}^{\Delta}$ to map the face stimuli in face-space representations and perform recognition tasks. Thus, our model provided d predictions for each test sample during each iteration. To gather a single prediction, we selected the most frequent class (mode) predicted by the face-space model for each test sample during each iteration, as per a majority voting approach. For each iteration, we then computed the overall recognition rate for the baseline approach and the Δ face-space in both facial expression and identity recognition conditions. This process led to 35 samples for each considered approach and task.

Results

The distribution of the sampled recognition rates was first assessed for normality using a D'Agostino's K-squared test (D'Agostino & Pearson, 1973) finding that the samples from both facial expression and identity tasks followed a normal distribution (p-values respectively 0.8571 and 0.1382). Thus, the effect between the baseline approach and our face-space model were evaluated by a Student's t-test (Keppel, 1991) at a significant level of $\alpha = 0.01$. The effect size was assessed by computing Cohen's d (Cohen, 1977).

The results for facial expression and identity recognition are shown in Figure 3a and Figure 3b respectively. From the plots, it is possible to see that the novel Δ face-space can facilitate both facial expression and identity recognition.

In addition, the t-tests rejected the null hypothesis in both facial expression (p-value=6.5e−19) and facial identity (p-



(a) (b)

Figure 3: Comparative analysis of the performance. (a,b) The recognition rates of the baseline approach and our face-space model respectively during facial expression and facial identity recognition tasks.

value=1.6e−6) recognition tasks. The computed effect size suggested a large effect for both the two tasks ($d = 3.03$ for facial expression recognition and $d = 0.98$ facial identity recognition). The statistics reached high powers (both > 0.98).

Potential Implications of the Hypothesis

Although we validated our hypothesis through computational simulations and it is not our aim to suggest that human brain implements the proposed face-space in this way, in this section we will discuss how these results can be of major importance for cognitive science community, at least by focusing at a functional level of analysis.

Modern literature in face perception studies widely suggest interactions between invariant and dynamic features of face stimuli. For instance, it has been shown that women and younger individuals appear to increase cues associated with happiness, whereas men and older people those of anger (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007) and studies in face processing broadly suggest that face stimuli can be plausibly represented in multidimensional norm-based spaces (Rhodes & Jeffery, 2006; Rhodes, Leopold, Calder, & Rhodes, 2011) and that invariant and dynamic codings of these spaces interact (A. J. Calder et al., 2001).

Interestingly, the proposed hypothesis well integrates with traditional understandings in face studies suggesting distinct routes processing invariant and dynamic features of the face, while still supporting more recent findings suggesting that representations of invariant and dynamic facial features partially overlap (Pell & Richards, 2013). In fact, Haxby, Hoffman, and Gobbini (2000) suggest that changeable aspects of the face (*i.e.* eye gaze, expression and lip movement) are processed in the Superior Temporal Sulcus (STS), whereas invariant aspects of the face necessary to classify the exhibited identity are processed in a distinct brain area, the Lateral Fusiform Gyrus (LFG). The STS presents neural connections with the amygdala and other brain areas usually associated with emotional processing capabilities (Adolphs, 2002) and interactions were observed between the STS and the LFG (Haxby et al., 2000). Recent neuroscience studies suggest that the STS is also related to mirroring mechanisms and imitative capabilities (Buxbaum, Shapiro, & Coslett, 2014) and Molenberghs, Brander, Mattingley, and Cunnington (2010) provided evidence suggesting that the role of the STS in imitation is not only to passively register observed biological motion, but rather to actively represent sensory-motor correspondences between one's actions and the actions of others. Therefore, the STS, assisted by putative emotional brain areas like the amygdala, can plausibly provide information necessary to interpret the observed facial expression, as suggested in this paper with the assumed system $\mathcal{M}otor$. This information, in turn, can be then used by the LFG to develop facial identity recognition capabilities, as proposed by the psychological face-space discussed in this paper.

Conclusions

We provided a new understanding of the face-space framework proposed by Valentine (1991) and able to realise a twofold structure encoding invariant and dynamic features of the face under shared codings and consequently facilitating facial expression and identity recognition capabilities. This face-space can develop by only interpreting motor behaviour exhibited by face stimuli encountered during development. We demonstrated the validity of our claim by providing compelling computational evidence and we discussed the potential implications of this new theoretical understanding in face perception and processing studies. Future works aim in extending the model with non-linear techniques and possibly include temporal features, while at the same time testing the theory by collecting human data from perceptual experiments.

References

Adolphs, R. (2002). Neural systems for recognizing emotion.

Current Opinion in Neurobiology, *12*(2), 169–177.

Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, *92*(2), 179.

Buxbaum, L. J., Shapiro, A. D., & Coslett, H. B. (2014). Critical brain regions for tool-related and imitative actions: a componential analysis. *Brain*.

Calder, A. (2011). *Oxford handbook of face perception*. Oxford University Press.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*(9), 1179–1208.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (revised ed.).* New York: Academic Press.

D'Agostino, R., & Pearson, E. (1973). Tests for departure from normality. empirical results for the distributions of b2 and b1. *Biometrika*, *60*(3), 613–622.

Gallese, V., & Caruana, F. (2016). Embodied simulation: beyond the expression/experience dualism of emotions. *Trends in Cognitive Sciences*.

Grossmann, T. (2015). The development of social brain functions in infancy. *Psychological Bulletin*, *141*(6), 1266.

Grossmann, T., & Vaish, A. (2009). Reading faces in infancy: developing a multi-level analysis of social stimulus. In T. Striano & V. Reid (Eds.), *Social cognition: Development, neuroscience and autism.* Oxford, UK: Blackwell Publishing.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Iacoboni, M. (2009). Do adolescents simulate? developmental studies of the human mirror neuron system. In T. Striano & V. Reid (Eds.), *Social cognition: Development, neuroscience and autism.* Oxford, UK: Blackwell Publishing.

Keppel, G. (1991). *Design and analysis: A researcher's handbook.* Prentice-Hall, Inc.

Le, Q. V., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G. S., … Ng, A. Y. (2013). Building high-level features using large scale unsupervised learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8595–8598).

Lee, K., Byatt, G., & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: Testing the face-space framework. *Psychological Science*, *11*(5), 379–385.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, 91–630.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 702–709.

Meltzoff, A. N., & Moore, M. K. (1992). Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behavior and Development*, *15*(4), 479–505.

Mobahi, H., Collobert, R., & Weston, J. (2009). Deep learning from temporal coherence in video. In *Proceedings of the 26th Annual International Conference on Machine Learning* (pp. 737–744).

Molenberghs, P., Brander, C., Mattingley, J. B., & Cunnington, R. (2010). The role of the superior temporal sulcus and the mirror neuron system in imitation. *Human Brain Mapping*, *31*(9), 1316–1326.

Ngo, T. T., Bellalij, M., & Saad, Y. (2012). The trace ratio optimization problem. *SIAM review*, *54*(3), 545–569.

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., … Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology*, *26*(10), 1334–1338.

Pell, P. J., & Richards, A. (2013). Overlapping facial expression representations are identity-dependent. *Vision Research*, *79*, 1–7.

Rhodes, G., Jaquet, E., Jeffery, L., Evangelista, E., Keane, J., & Calder, A. J. (2011). Sex-specific norms code face identity. *Journal of Vision*, *11*(1), 1.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, *46*(18), 2977–2987.

Rhodes, G., Leopold, D. A., Calder, A., & Rhodes, G. (2011). Adaptive norm-based coding of face identity. *The Oxford Handbook of Face Perception*, 263–286.

Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(6), 1113–1133.

Simion, F., & Di Giorgio, E. (2015). Face perception and processing in early infancy: inborn predispositions and developmental changes. *Frontiers in Psychology*, *6*.

Simpson, E. A., Maylott, S. E., Heimann, M., Subiaul, F., Paukner, A., Suomi, S. J., & Ferrari, P. F. (2016). Commentary on "Animal studies help clarify misunderstandings about neonatal imitation" by Keven and Akins.

Simpson, E. A., Murray, L., Paukner, A., & Ferrari, P. F. (2014). The mirror neuron system as revealed through neonatal imitation: presence from birth, predictive power and evidence of plasticity. *Philosophical Transactions of the Royal Society B*, *369*(1644).

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, *43*(2), 161–204.

Valentine, T., Lewis, M. B., & Hills, P. J. (2015). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, 1–24.

Vitale, J., Williams, M.-A., & Jonhston, B. (2016, August). The face-space duality hypothesis: a computational model. In *38th Annual Meeting of the Cognitive Science Society* (p. 514-519).