

# Stimuli selection procedure

Supplement materials of “I DARE: IULM Dataset of Affective Responses”

Marco Bilucaglia<sup>1,2,\*</sup>, Margherita Zito<sup>1,2</sup>, Alessandro Fici<sup>1,2</sup>, Chiara Casiraghi<sup>1,2</sup>,  
Fiamma Rivetti<sup>1</sup>, Mara Bellati<sup>1</sup>, and Vincenzo Russo<sup>1,2</sup>

<sup>1</sup>Behaviour and Brain Lab - Neuromarketing Research Center, Università IULM,  
Milan, Italy

<sup>2</sup>Department of Business, Law, Economics and Consumer Behaviour “Carlo A.  
Ricciardi”, Università IULM, Milan, Italy

\*Correspondence: marco.bilucaglia@studenti.iulm.it

We applied the following 7-steps semi-automatic stimuli selection procedure to IAPS (Lang et al., 2008) and OASIS (Kurdi et al., 2017), two datasets of realistic pictures that have been annotated in the valence and arousal dimensions. The annotations are provided as mean ( $m$ ) and standard deviation ( $s$ ) across the raters, for both the valence ( $m_v, s_v$ ) and arousal ( $m_a, s_a$ ).

Steps 1-5 of the procedure were based on the geometrical proprieties of the stimuli in the 2-dimensional Valence-Arousal space, that were considered either as points  $\mathbf{x} = (m_a, m_v)^T$ , or gaussian vectors  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , with  $\boldsymbol{\mu} = (m_v, m_a)^T$  and  $\boldsymbol{\Sigma} = \begin{pmatrix} s_v & 0 \\ 0 & s_a \end{pmatrix}$ . Steps 6-7 were based on the ratings provided by 9 independent judges.

## Step 1

From the annotations range  $R$ , we computed the central point  $\mathbf{c}$  and the peak-to-peak span  $p$  as, respectively:

$$\mathbf{c} \doteq (v_c, a_c)^T = \left( \min\{R\} + \frac{\max\{R\} - \min\{R\}}{2}, \min\{R\} + \frac{\max\{R\} - \min\{R\}}{2} \right)^T \quad (1)$$

$$p = \max\{R\} - \min\{R\} \quad (2)$$

## Step 2

We extracted the following subsets of points belonging to the four affective quadrants HVLA (High Valence High Arousal), HVLA (High Valence Low Arousal), LVHA (Low Valence High Arousal) and LVLA (Low Valence Low Arousal), excluding near-to-neutral points:

$$\text{HVHA} = \{\mathbf{x} \mid v > v_c + 0.1 \cdot p, a > a_c + 0.1 \cdot p\} \quad (3)$$

$$\text{HVLA} = \{\mathbf{x} \mid v > v_c + 0.1 \cdot p, a < a_c - 0.1 \cdot p\} \quad (4)$$

$$\text{LVHA} = \{\mathbf{x} \mid v < v_c - 0.1 \cdot p, a < a_c - 0.1 \cdot p\} \quad (5)$$

$$\text{LVLA} = \{\mathbf{x} \mid v < v_c - 0.1 \cdot p, a > a_c + 0.1 \cdot p\} \quad (6)$$

### Step 3

For each quadrant  $l = \{\text{HVHA}, \text{HVLA}, \text{LVHA}, \text{LVLA}\}$ , we computed gaussian centroid vectors  $\pi_l = \mathcal{N}(\mu_l, \Sigma_l)$ , where the mean vector and the covariance matrix are defined as, respectively:

$$\mu_l = \frac{1}{\#\mathcal{S}_l} \sum_{\mathbf{x} \in \mathcal{S}_l} \mu_{\mathbf{x}} \quad (7)$$

$$\Sigma_l = \frac{1}{\#\mathcal{S}_l} \sum_{\mathbf{x} \in \mathcal{S}_l} \Sigma_{\mathbf{x}} \quad (8)$$

$\mu_{\mathbf{x}}$  and  $\Sigma_{\mathbf{x}}$  are, respectively, the mean vector and covariance matrix of  $\mathbf{x}$ ,  $\mathcal{S}_l$  is the set of  $\mathbf{x}$  belonging to the quadrant  $l$  and  $\#\mathcal{S}_l$  is the cardinality of  $\mathcal{S}_l$ .

### Step 4

For each vector  $\mathbf{x} = \mathcal{N}(\mu, \Sigma) \in \mathcal{S}_l$  we computed its distribution similarity with the centroid vector  $\pi_l$  by means of the Bhattacharyya distance  $d(\mathbf{x}, \pi_l)$ , defined as (Choi and Lee, 2003):

$$d(\mathbf{x}, \pi_l) = \frac{1}{8}(\mu - \mu_l)^T (\Sigma')^{-1} (\mu - \mu_l) + \frac{1}{2} \ln \left( \frac{\det(\Sigma')}{\sqrt{\det(\Sigma) + \det(\Sigma_l)}} \right) \quad (9)$$

where  $\Sigma' = (\Sigma + \Sigma_l) / 2$  and  $\det(\cdot)$  is the determinant.

### Step 5

From each quadrant  $l$  we extracted the 6 stimuli with the highest Bhattacharyya distances.

### Step 6

Nine independent raters selected, from each set of 6 stimuli, a subset of 4.

### Step 7

We selected the 4 stimuli that obtained the highest percentage agreement (McHugh, 2012) across the raters. In case of parity, we considered the highest distance from the centroid.

## References

- Choi, E. and Lee, C. (2003). Feature extraction based on the Bhattacharyya distance. *Pattern Recognition*, 36(8):1703–1709.
- Kurdi, B., Lozano, S., and Banaji, M. R. (2017). Introducing the open affective standardized image set (OASIS). *Behavior research methods*, 49(2):457–470.
- Lang, P., Bradley, M., and Cuthbert, B. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical report, University of Florida, Gainesville.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.