Computing a unique neural fingerprint of human bodily actions and expressions

Vojtěch Smekal[1], Marta Poyo Solanas[1], Beatrice de Gelder[1,2]

[1]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Limburg 6200MD, The Netherlands
[2]Department of Computer Science, University College London, London, WC1E 6BT, UK

Goal

Discover feature combinations representative of different action classes and their neural representations

Background

Mid-level features of bodies and body expressions have been proposed as behaviourally relevant information coded in the brain (e.g. limb contraction for fear coded in fear related areas - Poyo Solanas et al., 2020), bridging the gap between low-level features of visual input and high-level cognitive labels requiring conscious processing (de Gelder & Poyo Solanas, 2021). This mid-level feature approach can help us understand the exact functions of brain regions identified as being involved in body perception. For instance, the extrastriate body area (EBA) and fusiform body area (FBA) have been shown to be involved in the processing of whole bodies and body parts (Peelen & Downing, 2005). The posterior superior temporal sulcus (pSTS) is involved in the perception of biological action, showing greater activation for dynamic versus static social stimuli (Landsiedel, Daughters, Downing & Koldewyn, 2022). The posterior parietal cortex is thought to have functional motor maps for action classes grouped based on their motor goals (Orban, Lanzilotto & Bonini, 2021).

Stimuli

Stimuli: 6 actors performing 6 different actions (3 emotional and 3 neutral)

- Defensive, Greeting, Raging, Grooming, Peeling a banana, Searching/Foraging

Stimuli presented in 3 different conditions:

- Normal video: 1 second, 50-frame clip of an action performed by one person, made grayscale and controlled so direction of motion is consistent across actors
- Static image: single frames extracted from each video using the frame-by- frame algorithm (Bockes & Vrabie, 2021), which chooses the frame with the highest softmax value for a given ground-truth action label
- Frame-scrambled video: the order of the frames of the original video was pseudorandomised to create a new 1 second video; each video was divided into 5 parts of 10 frames each and the order of these parts wass then shuffled

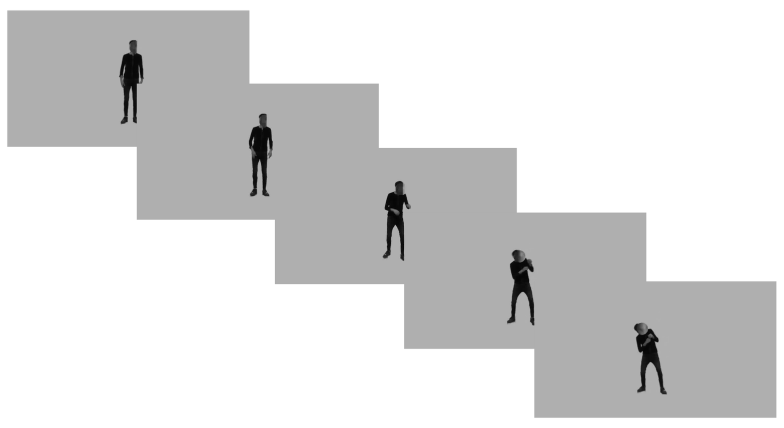

Figure 1. A set of example frames from a normal video.

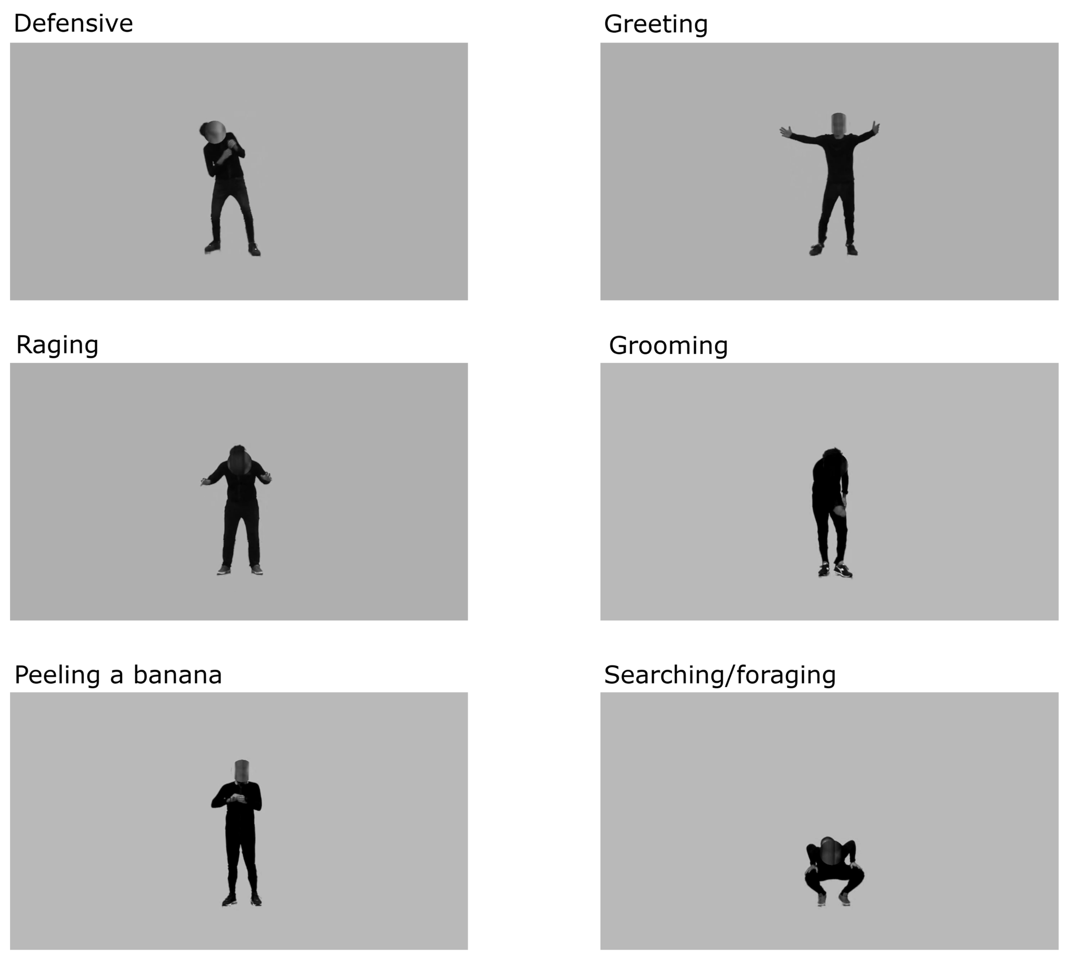

Figure 2. Example frames from each of the six action categories.

Body Action Analysis

- Each video stimulus was analysed using the OpenPose algorithm (Cao et al., 2017) to extract the position of 25 key points for each frame of a video
- The key points were manually corrected and normalised, so each was located in the same starting space

Hierarchical Clustering

Hierarchical clustering was then conductted on the key points, in order to group together key points which were often correlated in their movement, creating feature clusters for describing each action category. The hierarchical clustering was combined with the elbow method to determine the optimal number of clusters. The elbow methods is a form of otpmization, which chooses a point where adding an additional cluster would not be worth the cost to the complexity. This resulted in 5 feature clusters to describe each action with:

- right arm, head and upper torso, legs, left hand, lower torso and left elbow

Principal Component Analysis

Principal component analysis was also run on the key point data, which identified four principal components. These can be roughly described as:

- legs, head and upper torso, torso and left foot, arms

Conclusions

Both the hierarchical clustering and the PCA allow us to describe the actions performed by the actors in terms of a small number of clusters/components, which can then be used in an RSA with the brain activation data.

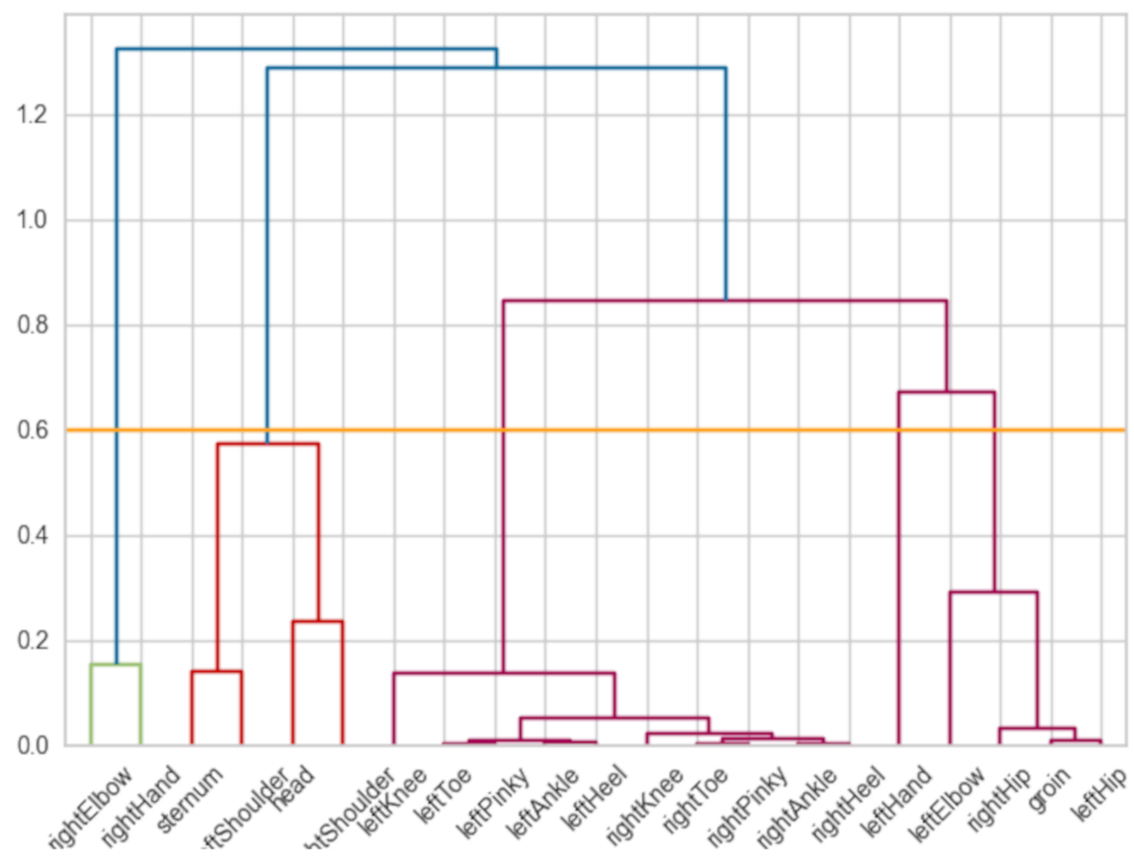

Figure 3. The results of the hierarchical clustering, with the yellow horizontal line indicating the appropriate cut-off point based on the elbow method.

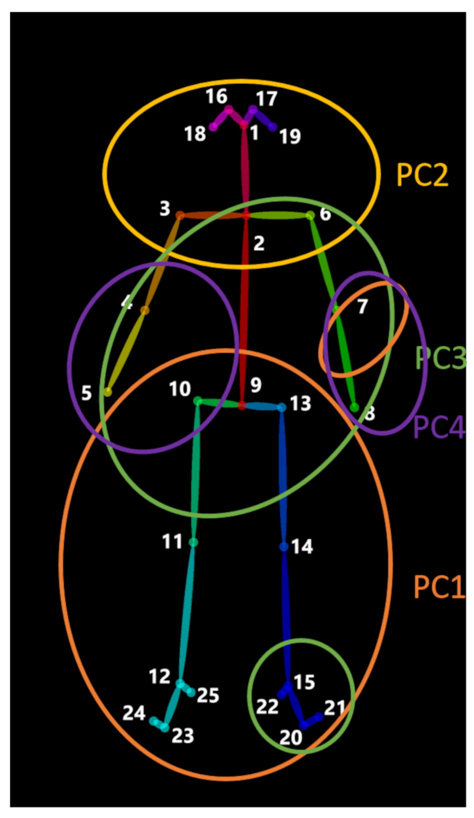

Figure 4. The results of the PCA analysis.

fMRI Procedure

- 17 participants (9 female)
- 8 scanning runs consisting of 114 stimuli each (6 actions x 6 actors x 3 conditions + 6 catch trials)
- 2 runs of functional localizer (bodies, faces, objects)
- When the fixation cross changed shape to a circle, the participant has to press (1) if the current stimulus is identical to the previous one or (2) if it is different to the preceding stimulus

- 3T Siemens Prisma MRI Scanner at Scannexus, Maastricht University
- Functional Imaging: accelerated multi-band 4 T2*-weighted sequence (TR = 1300 ms, TE = 23 ms, 56 slices, 2 mm isotropic)
- Anatomical Imaging: MPRAGE T1-weighted sequence, slice thickness = 1 mm, in-plane resolution = 1 x 1 mm

Next steps

The brain activation evoked by the body movement stimuli will be investigated using GLM approaches with defined contrasts (e.g. actions, conditions, emotionality), as well as using RSA combining brain activation with the body action analysis.

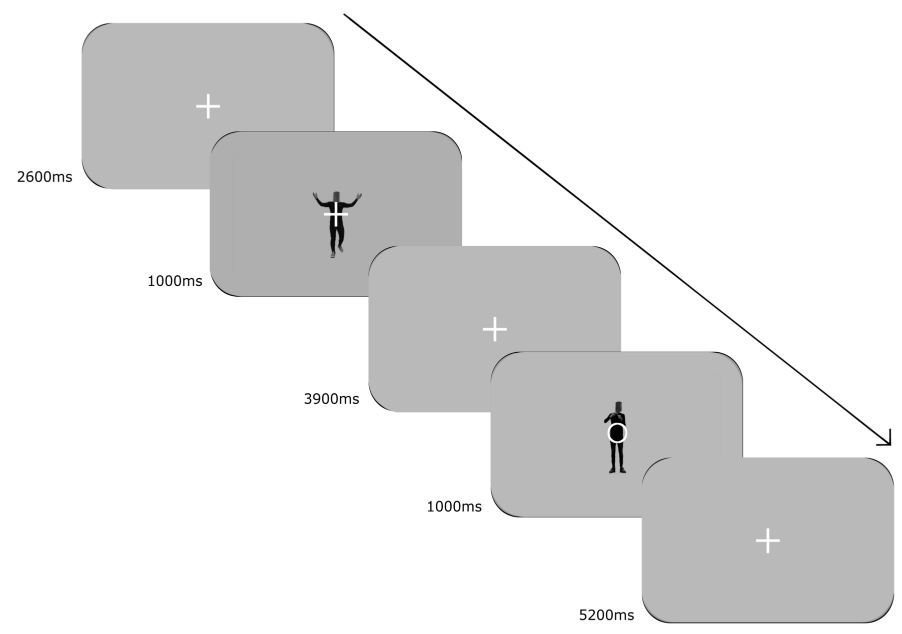

Figure 5. An example fMRI trial procedure, including an example catch trial, where the participant must make a button response when the fixation ross becomes a circle - pressing '1' if the stmulus is the same as the previous or '2'if it is different.

References

- Cao, Z., Simon, T., Wei, S-E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.* 7291-7299.
- de Gelder, B., & Poyo Solanas, M. (2021). A computational neuroethology perspective on body and expression perception. *Trends in cognitive sciences*, *25*(9), 744–756.https://doi.org/10.1016/j.tics.2021.05.010
- Landsiedel, J., Daughters, K., Downing, P.E., & Koldewyn, K. (2022). The role of motion in the neural representation of social interactions in the posterior temporal cortex.*NeuroImage*, *262*.
- Orban, G. A., Lanzilotto, M., & Bonini, L. (2021). From Observed Action Identity to Social Affordances. *Trends in Cognitive Sciences*, *25*(6), 493–505. https://doi.org/https://doi.org/10.1016/j.tics.2021.02.012
- Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of neurophysiology*, *93*(1), 603–608. https://doi.org/10.1152/jn.00513.2004
- Poyo Solanas, M., Vaessen, M., & de Gelder, B. (2020). The role of computational and subjective featrues in emotional body expressions. *Sci Rep, 10.* 1-13.

Funding

This work wsa supported by the ERC Synergy grant (Grant agreement 856495; Relevance), by the Future and Emerging Technologies (FET) Proactive program H2020-EU.1.2.2 (Grant agreement 824160; EnTimeMent) and by the Industrial Leadership Program H2020-EU.1.2.2 (Grant agreement 825079; MindSpaces).