# Dynamic facial expressions are not necessarily processed holistically

Eva R. M. Joosten†, Martin A. Giese†

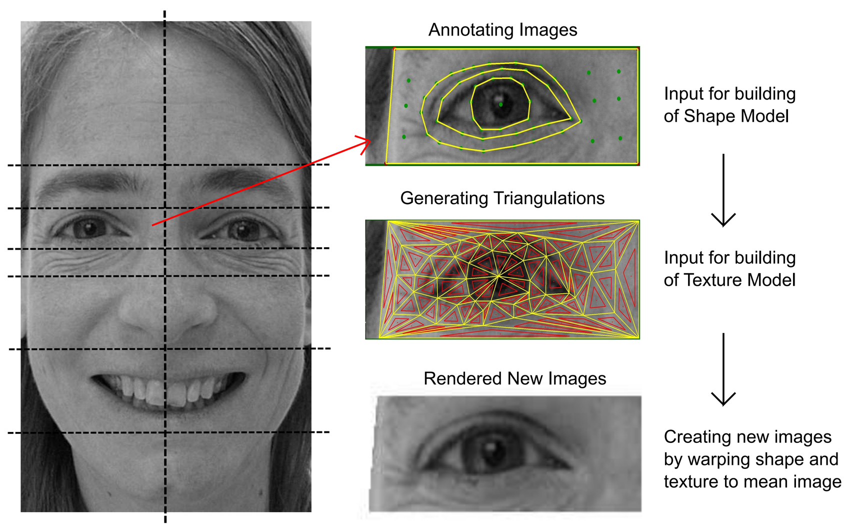
†Section for Computational Sensomotorics., Dept. of Cognitive Neurology, University Clinic, CIN, HIH, University of Tübingen.

eva.joosten@uni-tuebingen.de, martin.giese@uni-tuebingen.de

## Introduction

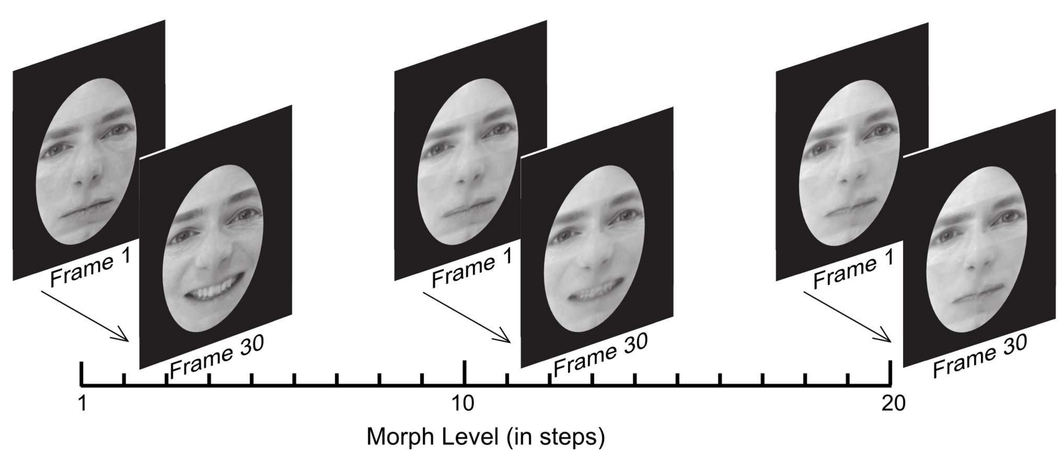
Face perception studies support a holistic/part-based model of face perception, where holistic and part-based processing make parallel and separable contributions to face perception [Piepers & Robbins 2012]. However, it remains unclear which exact aspects are included in holistic processing. Many studies have concluded that facial expressions are processed holistically [e.g. [White 1999], [?]], but most of them have used static stimuli, opposed to the fact that natural facial expressions are highly dynamical. Using a novel technique to control the dynamics of individual face components separately, we studied the integration of dynamic facial components in expression recognition. To parameterize spatial and temporal variations by low dimensional vectors of facial components, our model exploits Active Appearance Models [Cootes et al. 1998]. Via staircase procedures, we varied the spatial and temporal properties of information in the different components. Participants detected dynamic emotional expressions ranging from no ambiguity at all, towards a static image with neutral expression (Exp. 1). Participants also detected asynchrony between facial components (Exp. 2). We determined separate thresholds for different components and found that participants were equally accurate in detecting emotional expressions as long as the eyes were informative. We also found that participants' temporal resolution is more sensitive when top halfs of the faces are being slowed down, even when faces are inverted.

## Stimuli

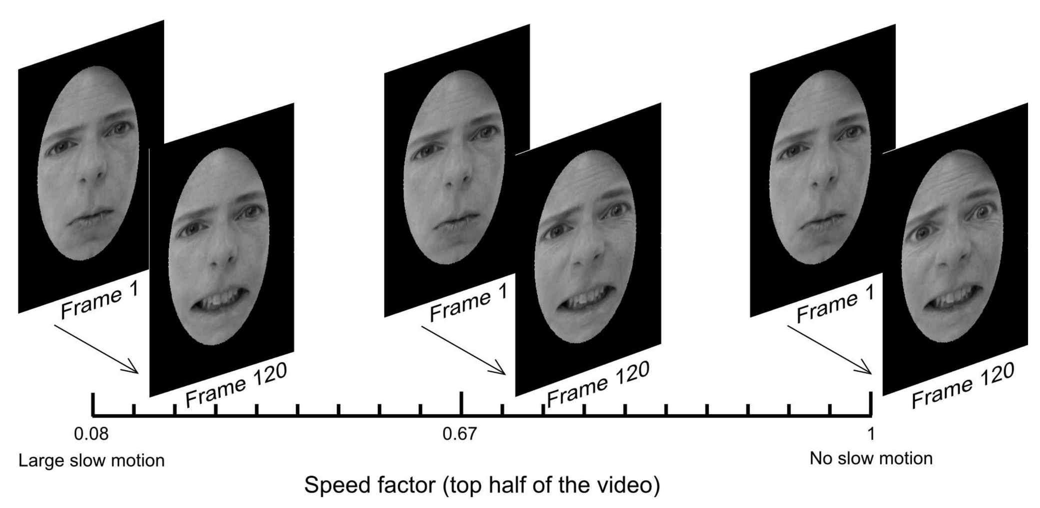



**Exp. 1** Stimuli were modeled by Active Appearance Models [Cootes et al. 1998] that were trained separately on different facial components. This computer vision algorithm fits statistical models of object shapes and appearances to new images. To train the models, we provided 90 HD video frames of emotional expressions and computed coordinates of landmarks for each frame. The models for individual components determined the shape and texture in terms of low dimensional vectors.

Stimuli (500 ms) were presented with a 2-up-1-down staircase procedure. Morph levels were ranging from 1 (fully expressive dynamic face) to 20 (neutral face). Only certain parts of the face carried information; one full face condition (MNE) and 6 partial face conditions: eyes (E), nose (N), mouth (M), nose + eyes (NE), mouth + eyes (ME) and mouth + nose (MN).



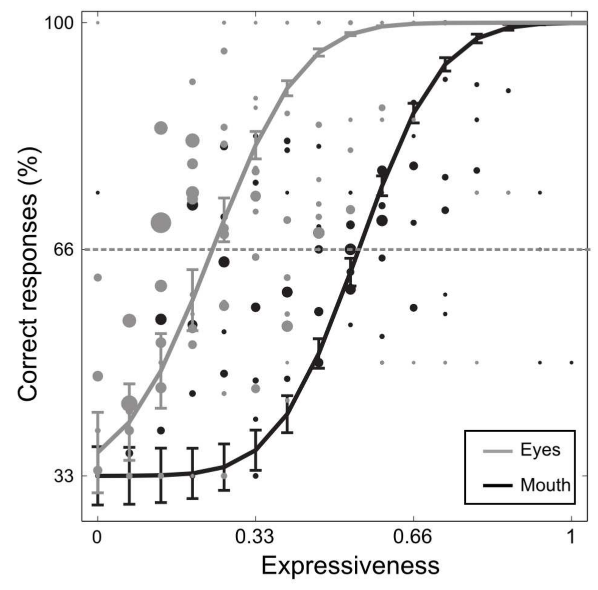

**Exp. 2** Asynchrony between facial halfs was created by reducing the speed factor of one part as opposed to the other. Stimuli (4 s) consisted of 240 time warped frames. The figure below displays frame 120 to indicate the differences in optic flow of the top half as opposed to the bottom half. When there is no slow motion (right), top and bottom halfs play simultaneous at 30 FPS. When the speed factor is reduced, top and bottom halfs get more asynchronous. Via staircase procedures, stimuli (4 s) were interleaved presented in four conditions; slow motion in top or bottom halfs and an upright versus inverted presentation of the stimuli.




## Results I

**Task** After each trial, participants (5 male, 5 female) were asked to press "F" for fearful, "H" for happy and "A" for angry expressions. The number of collected trials was 35K (1692±646, mean±SD across 21 conditions (3 emotions x 7 facial components).

**Results** Threshold curves including difference thresholds ($\mu$) and standard deviations ($\sigma$) were calculated for the 21 conditions. We tested for differences between the threshold curves for the different conditions using Kolmogorov-Smirnov tests. In comparison with other components, we found significantly lower thresholds for the conditions where the emotion specific information was attributed to the eyes (all $p$'s $< 0.05$). The thresholds for the perception of facial expressions were equal for holistic faces and combinations of spatial components, as long as the eyes were present (all $p$'s $> 0.14$).





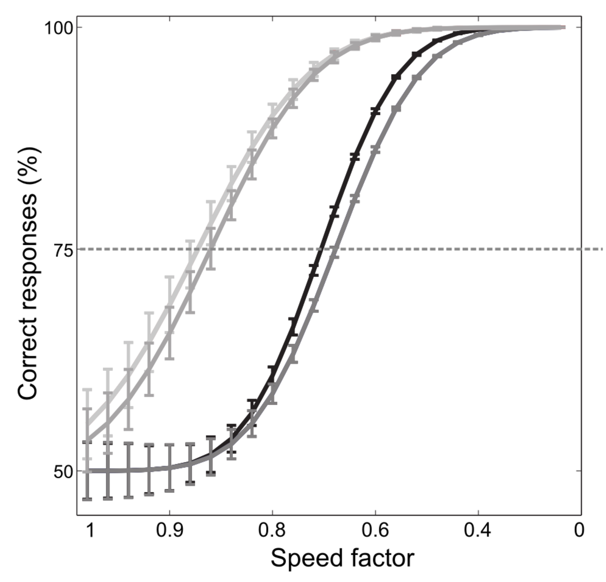
| | M | N | MN | E | ME | NE | MNE |
|---|---|---|---|---|---|---|---|
| $\mu_{fear}$ | 0.45 | 0.41 | 0.46 | 0.19 | 0.16 | 0.16 | 0.16 |
| $\mu_{happy}$ | 0.39 | 0.12 | 0.47 | 0.28 | 0.33 | 0.34 | 0.36 |
| $\mu_{anger}$ | 0.45 | 0.19 | 0.47 | 0.2 | 0.31 | 0.27 | 0.23 |
| $\sigma_{fear}$ | 0.24 | 0.35 | 0.25 | 0.14 | 0.12 | 0.15 | 0.13 |
| $\sigma_{happy}$ | 0.10 | 0.33 | 0.12 | 0.18 | 0.14 | 0.26 | 0.12 |
| $\sigma_{anger}$ | 0.12 | 0.15 | 0.11 | 0.12 | 0.09 | 0.12 | 0.11 |

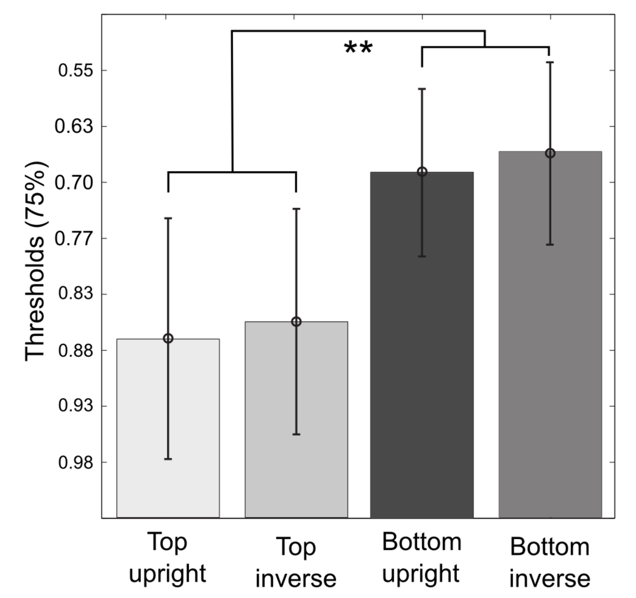
⇒ Not all facial components contribute to the perception of dynamic expressions. It is the information present in the eyes that determines perceptual accuracy.

## Results II

**Task** Participants (5 male, 5 female) were asked to indicate whether the part containing the slow motion (d.i. the reduced speed factor) was in the top or bottom half of the video. Stimuli were presented upright and inverted. The number of collected trials was 7532 (1883±80, mean±SD across 4 conditions (2 halfs x 2 orientations).

**Results** Threshold curves including difference thresholds ($\mu$) and standard deviations ($\sigma$) were calculated for the 4 conditions.




⇒ Participants are significantly (all $p$'s $< 0.001$) more sensitive to asynchrony when the reduced speed factor is presented in the top half of the video. The effect remains when stimuli are inverted

## Discussion

In social interactions, it often happens that facial expressions are partially occluded (e.g. due to angle differences as opposed to the observer) or that certain parts of the face express ambiguous emotional expressions. Yet, often we are still able to describe the expressions.

This research aimed to design an experimental setup with natural dynamic faces that are partially occluded but still embedded in a natural context (e.g. not by foreground masking). Participants are able to detect dynamic emotional expressions equally accurately as long as the eyes are informative. So far, this effect is not shown with dynamic stimuli and contrasts with the classical effect where face recognition is impaired when only one component of the face is presented.

With respect to temporal asynchrony, participants are more sensitive to asynchrony when the speed factor is reduced in the top as opposed to the bottom half of the video. This effect was preserved for inverted stimuli, abstract faces and objects. Differences in visual processing of inverted as opposed to upright faces is an indicator for holistic processing ([Piepers & Robbins 2012]). We did not find indications of a holistic processing of dynamic emotional expressions.

## References

[Blanz & Vetter 1999] V. Blanz and T. Vetter, *A morphable model for the synthesis of 3D faces*, Siggraph 99 Conference Proceedings (1999), 187–194.

[Cootes et al. 1998] T. F. Cootes, G. J. Edwards and C. J. Taylor *Active appearance models*, IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (1998), no. 6, 681.

[Piepers & Robbins 2012] D. W. Piepers and R. A. Robbins *A review and clarification of the terms "holistic," "configural," and "relational" in the face perception literature*, Front Psychol **3** (2012), 559.

[White 1999] M. White *Representation of facial expressions of emotion*, Am J Psychol **112** (1999), no. 3, 371-381.

[Tanaka et al. 2012] . W. Tanaka, M. D Kaiser, S. Butler and R. Le Grand *Mixed emotions: holistic and analytic perception of facial expressions*, Cogn Emot **26** (2012), no. 6, 961-977

## Acknowledgements

This work was supported by the EU Commission, Fp 7-PEOPLE-2011-ITN(Marie Curie): ABC PITN-GA-011-290011, 7th Framework Programme: EC FP7-ICT-249858 TANGO, EC FP7-ICT-248311 AMARSi, The Human Brain Project, Deutsche Forschungsgemeinschaft: DFG GI 305/4-1, DFG GZ: KA 1258/15-1.