View dependence of complex versus simple facial motions

Christian Wallraven* Douglas W. Cunningham Martin Breidt Heinrich H. Bülthoff
Max Planck Institute for Biological Cybernetics, Tübingen, Germany

1 Introduction

In this study we investigate the *viewpoint dependency* of complex facial expressions versus simple facial motions (so-called "action units" [Ekman and Friesen 1978]). The results not only shed light on the cognitive processes underlying the processing of complex and simple facial motion for expression recognition, but also suggest ways how to incorporate these results into computer graphics and computer animation [Breidt et al. 2003]. For example, expression recognition might be highly viewpoint dependent making it difficult to recognize expressions from the side. As a direct consequence, modeling of expressions would then require only the frontal views to "look good", i.e., it would in principle be unnecessary to attempt detailed 3D modeling of expressions. If, however, recognition of expressions were view-invariant, then modeling would have to provide a faithful 3D rendering of facial expressions.

For the psychophysical investigation of view dependency of facial expressions we used the MPI Tübingen Facial Expression Database[1]. This database was created to provide a set of high-resolution, time-synchronized video sequences of facial expressions from several viewpoints. It should enable researchers to address a wide variety of scientific questions in cognitive research (such as done in this paper), computer vision, human-computer interaction, communication research and computer graphics.

2 Experiment and results

From the video database we first extracted 14 action units, which included only internal motions of the face (no rigid head motion). In addition to these action unit sequences, 8 complex facial expressions were taken from the database. All sequences were recorded from four viewpoints spanning a total of 68°. Ten participants took part in the experiment, which consisted of a 22 alternative-forced choice task in which participants were instructed to view a looping video sequence and to indicate as quickly and accurately as possible which of the 14 action units or 8 facial expressions was depicted in the sequence (based on a table of names of both expressions and action units). Dependent variables in this experiment were reaction time (RT) and recognition accuracy. Statistical analysis was done using a multivariate ANOVA with factors "expression type" and "viewpoint" based on "reaction time" and "recognition accuracy".

Participants had an average recognition accuracy of 88.6%, showing that the task was not too hard. Interestingly, an analysis of the confusion matrix showed that expressions were never confused with action units and vice versa, which demonstrates a clear semantic separation of simple from complex facial motions. The ANOVA

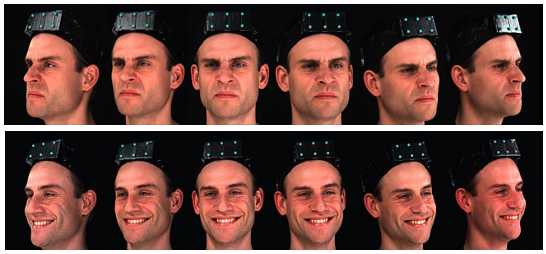

Figure 1: Six views of the peak expression from the video sequence of a) an action unit (au09, "nose wrinkler") and b) "happy"

revealed at best *marginally significant effects* of viewpoint (p=0.07, n.s.) or type of expression (p=0.07 n.s.) on recognition accuracy. These effects, however, were mainly caused by three action units, which were often misclassified. Indeed, an additional analysis without these action units found no effects of either viewpoint (p=0.20, n.s.) or type of expression (p=0.31,n.s.). With an average reaction time of 2.8s, participants responded after one repetition of the sequence. The analysis of reaction times again revealed *no effects* of either viewpoint (p=0.54, n.s.) or type of expression (p=0.71 n.s.).

One of the possible effects of view-dependent recognition could be that the recognition time varies with viewpoint: it might be more time-consuming to extract facial motion information from side views. Recognition accuracy might also be affected by viewpoint: facial motion might be more ambiguous from the side than from the front, for example. The experimental results, however, showed no clear effects of viewpoint on either factors. It thus seems that humans are able to recognize facial motions in a largely *viewpoint invariant* manner (at least within the viewing range covered in this experiment), which supports the theoretical model of face recognition by [O'Toole et al. 2003]. In addition, our results suggest that in order to be recognized, computer generated facial expressions should "look good" from all viewpoints.

The fact that we found no differential effects of action units and expressions sheds further light on processing strategies of expressions. First, untrained participants were able to recognize action units with a surprisingly high accuracy. Second, recognition performance of full expressions *cannot* be explained by simply adding the observed recognition performance of their constituent action units (for example, "sad" can be constructed by three simple action units). It thus seems that "the whole is more than the sum of its parts". Future experiments will need to clarify exactly how big this advantage is and how a complex of action units can be used to model recognition performance. This in turn will have an impact on how to best generate recognizable and believable facial expressions [Cunningham et al. 2003; Breidt et al. 2003] .

References

BREIDT, M., WALLRAVEN, C., CUNNINGHAM, D., AND BÜLTHOFF, H. 2003. Facial animation based on 3d scans and motion capture. In *Siggraph'03 Sketches and Applications*.

CUNNINGHAM, D., BREIDT, M., KLEINER, M., WALLRAVEN, C., AND BÜLTHOFF, H. How believable are real faces: Towards a perceptual basis for conversational animation. In *Proc. of Computer Animation and Social Agents 2003*.

EKMAN, P., AND FRIESEN, W. 1978. *Facial Action Coding System (FACS)*. Consulting Psychology Press.

O'TOOLE, A., ROARK, D., AND ABDI, H. 2003. Recognizing moving faces: a psychological and neural synthesis. *Trends in Cognitive Sciences 6*, 6.

*firstname.lastname@tuebingen.mpg.de

[1] http://www.kyb.mpg.de/~mbreidt/au_videos