Decoupling facial expressions and head motions in complex emotions

Andra Adams, Marwa Mahmoud, Tadas Baltrušaitis and Peter Robinson
Computer Laboratory
University of Cambridge, Cambridge, UK
{Andra.Adams, Marwa.Mahmoud, Tadas.Baltrusaitis, Peter.Robinson}@cl.cam.ac.uk

***Abstract*—Perception of emotion through facial expressions and head motion is of interest to both psychology and affective computing researchers. However, very little is known about the importance of each modality individually, as they are often treated together rather than separately. We present a study which isolates the effect of head motion from facial expression in the perception of complex emotions in videos. We demonstrate that head motions carry emotional information that is complementary rather than redundant to the emotion content in facial expressions. Finally, we show that emotional expressivity in head motion is not limited to nods and shakes and that additional gestures (such as head tilts, raises and general amount of motion) could be beneficial to automated recognition systems.**

Index Terms*—affective computing*; *emotion recognition

I. Introduction

Humans are adept both at expressing themselves and interpreting others through the use of non-verbal cues [1]. Facial expressions, eye gaze and the head are important modalities for non-verbal communication. They communicate rich social content, such as turn-taking [2], intention [3], and emotion [4].

Emotional expression through these modalities has been an area of interest in both psychology and affective computing [4]–[6]. However, little is known about the interaction between modalities and the importance of each individual modality, as they are often treated together instead of separately. Our work attempts to fill this knowledge gap by isolating the influence of the face and of the head in the perception of complex emotions.

Many prior emotion perception studies used only static images rather than dynamic videos for their emotion perception tasks, which precludes any movement-related analysis. Such analysis is important as the speed and sequence (i.e. specific gestures) of movement carry important emotional content [7]–[10]. Moreover, among the studies which did use dynamic stimuli for head motions, there is a lack of research into gestures beyond nods and shakes. In our work we use dynamic video stimuli to analyze the emotion recognition rates of isolated facial expressions and head motions.

In affective computing, most of the research in vision-based emotion recognition has centered around facial expressions [6], [11]. Some work combines head pose, eye gaze and facial expressions [12], [13] but few distinguish between the contributions of each modality in identifying emotion. Better understanding of the importance of the head in the perception of dynamic complex emotions would allow us to build better affect recognition systems by guiding the selection of features and algorithms.

We present a study which isolates the effect of head motion from facial expression in the perception of complex emotions in acted videos. We give a quantitative analysis of emotion recognition rates for videos with facial expressions only and videos with head motions only, as compared to the recognition rates of the same videos with both modalities together as well as shoulder movements. See Figure 1 for sample still-frames of each of the three types of stimuli used in this study.

We demonstrate quantitatively that head motions carry emotional information that is independent of the emotion content in facial expressions, and that these head gestures extend beyond nods and shakes.

II. Previous Work

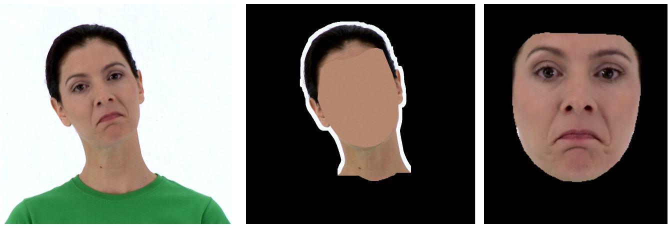
The face is one of the most important channels of nonverbal communication. Facial expressions figure prominently in research on almost every aspect of emotion [5]. However, head poses and head gestures also play an important role in emotion expression and perception.

Particular facial expressions in combinations with head poses have been associated with various emotional states [14]–[16]. The expressions of pride and shame include both facial expressions and changes in head posture [15]. Embarrassment, shame and guilt have associated face, head and body movements [16]. Anxiety can be detected from facial displays [17].

Angular displacement and angular velocity of the head in emotional dyadic scenarios have been correlated with broad concepts of "Conflict" and "Normal" [8]. Newborns have been found to imitate head gestures [18], pianists use head movements to convey subtle emotions [19], and 100 types of nods have been found in political debate footage [20].

However, the majority of previous work has ignored the confounding factors of facial expression and head motion, either by disregarding head motion entirely or by only ever analyzing both modalities together.

A few exceptions [21]–[23] have attempted to control for the influence of each modality, and have demonstrated the importance of the head in emotion perception. Hess et al. demonstrated that head position strongly influences reactions to anger and fear [23]. Direct anger expressions were more accurately decoded than averted anger expressions created by modifying head pose. Mignault et al. found that a raised



(a) Upper body (b) Head (c) Face

Fig. 1. Example still-frames from the three types of video stimuli: (a) upper body (head, face, shoulders), (b) isolated head, and (c) isolated facial area.

head is associated with happiness and 'superiority emotions' (e.g. pride, contempt), and a bowed head is associated with 'inferiority emotions' (e.g. guilt, humiliation) [21]. Tracy and Robins showed that the head raised 15–20 degrees encourages the recognition of pride [22].

However, none of these three studies completely decoupled the head or face from other modalities. They did not mask co-occuring facial expressions, yet even neutral expressions can carry emotional information [24]. Our work attempts to isolate the effects of the face (expression with eye gaze) and of the head in the perception of complex emotions.

III. Methodology

A. Dataset

The dataset of videos used is a subset of the EU-Emotion Stimulus Set (EESS), recently collected and validated as part of the ASC-Inclusion project [25].

Along with other multi-modal data, the EESS contains video of acted facial expressions and head motions. Each video is 2–17 seconds in length and has been labelled with *Neutral* or one of 17 emotion categories.

The EESS videos were each labelled by 57–580 people using a six-option forced-choice format. The six choices were: the target emotion that the actor was intending to express, four control emotions (selected per-emotion based on similarity scores between emotion categories [25]), and "None of the above" to prevent artifactual agreement [26]. Table I shows the specific foils assigned to each target emotion category.

In this study, we used only the 181 videos which met the validation requirement of having the target emotion chosen by at least 50% of the raters and no foil chosen by more than 25% of the raters [25].

B. Video Pre-Processing

1) Isolating head motions: To isolate the heads of the actors, each video was pre-processed to remove faces and shoulders. The pipeline is shown in Figure 2.

The Cambridge face tracker [27] was used to extract 68 facial landmarks from each video frame. The eyebrow region landmarks were shifted upwards by 6% of the image's height to mask forehead wrinkles caused by eyebrow movements.

Rather than using a neutral facial expression, we chose to use an opaque mask covering the face, as even neutral expressions can communicate emotional information [24].

TABLE I

Foils assigned to each emotion category in the EU-Emotion Stimulus Set [25]

Target	**Foil #1**	**Foil #2**	**Foil #3**	**Foil #4**
Afraid	Ashamed	Unfriendly	Disappointed	Kind
Angry	Jealous	Disgusted	Surprised	Happy
Ashamed	Disappointed	Worried	Unfriendly	Proud
Bored	Frustrated	Sad	Hurt	Excited
Disappointed	Worried	Bored	Afraid	Joking
Disgusted	Afraid	Frustrated	Sad	Interested
Excited	Interested	Joking	Hurt	Bored
Frustrated	Sad	Jealous	Sneaky	Kind
Happy	Interested	Surprised	Bored	Angry
Hurt	Worried	Unfriendly	Surprised	Happy
Interested	Excited	Proud	Joking	Disappointed
Joking	Kind	Interested	Proud	Angry
Neutral	Bored	Kind	Surprised	Frustrated
Proud	Excited	Interested	Kind	Afraid
Sad	Afraid	Jealous	Disgusted	Proud
Sneaky	Angry	Disappointed	Ashamed	Kind
Surprised	Happy	Joking	Worried	Bored
Worried	Angry	Disappointed	Disgusted	Happy

Thus a convex hull of the facial landmarks was created to mask the face. The colour and texture of the mask was chosen to match the skin colour of the actor.

Finally, hue, saturation and value (HSV) thresholding was used to detect and remove the white background and the actor's shoulders from the videos.

Of the 181 videos, 13 did not have adequate landmark detection throughout the video which caused parts of the facial expression to be visible even after the face masking. Thus only 168 head-only videos were acceptable as stimuli.

2) Isolating facial expressions: We processed each of the videos to isolate the facial expression (with eye gaze) and remove the effects of head orientation and head motion. The pipeline used to achieve this is illustrated in Figure 3.

Similar to the head isolation, we extracted the 3D location estimates of 68 facial landmarks together with the head orientation from each video frame. To capture any emotional information in the forehead region, we shifted the eyebrow landmarks upwards. We then rotated the facial landmarks to a frontal head pose to remove all head rotation. We used piece-wise affine warping between the original image and the resulting frontal landmarks to generate the isolated face image.

Our method for isolating the face worked in most cases, but there were sources of error which lead to some of the videos not being included. First, tracking was not always successful, resulting in incorrect facial alignment. Second, when the head pose is extreme the face is self-occluded meaning that piece-wise affine warping did not produce correct results.

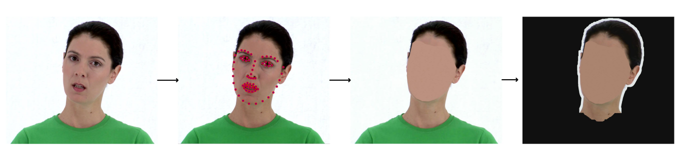


Fig. 2. Isolating the head from each video via landmark detection of 68 facial features. First the face region is replaced with an opaque skin colour. Next the shoulders are removed with Hue-Saturation-Value thresholding.

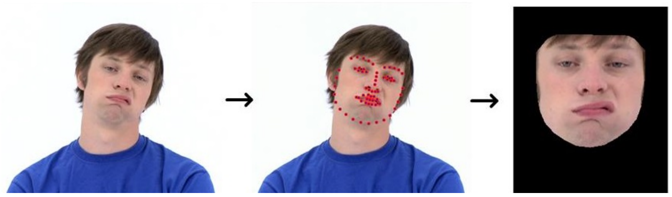


Fig. 3. Isolating the face from the emotional stimulus videos. First the facial features are detected in the video, followed by feature alignment into frontal head orientation using piece-wise affine warping of face triangles.

Of the 181 videos, 25 had incorrect alignment or self-occlusion and thus only 156 face-only videos were acceptable.

3) Final video selection: Of the 181 videos, 168 were adequate for head-only and 156 were adequate for face-only. The intersection of these two sets, and thus the final set of videos used in the surveys, was 153 videos.

Figure 1 shows sample still-frames from the three types of video stimuli: (a) Upper Body including face, head and shoulders, (b) the isolated Head, and (c) the isolated Face.

C. Crowdsourcing Interface

A simple web interface was developed to collect labels for the head-only and face-only videos. A video was randomly selected from the set of 153 and displayed to the participant. The participant was asked to select the label which best described what the person in the video was expressing. The same sets of foils from the original experiment (see Table I) were given as the possible responses, in randomized order. The video could be re-watched as many times as necessary.

Participants were recruited via social media and university mailing lists. The incentive offered was entry in a raffle for any participant who labelled more than 50 videos. The labelling period was approximately one week.

IV. Results

A. Inter-Rater Agreement

We collected a total of 3071 responses (123 raters) for the head-only videos and 3296 (81 raters) for the face-only videos. Since the videos were randomized, we did not get the same number of labels for each video. For the head-only videos, we had a mean of 20 labels per video (min=11, max=32, SD=4). For the face-only videos, we had a mean of 22 labels per video (min=13, max=34, SD=3.2).

To have confidence in the labels obtained from crowdsourcing, we evaluated inter-rater agreement using Krippendorff's Alpha measure [28], as it handles multiple raters and missing data. We obtained Alpha values of 0.37 for head-only videos and 0.49 for face-only videos, indicating fair and moderate agreement respectively. These levels of agreement rule out the possibility of agreement by chance, and are similar to those reported in other emotion perception studies [29].

B. Recognition Rates Across Modalities

Using the same validation criteria as the original study, head-only and face-only videos were considered to be reasonable portrayals of a particular emotion if at least 50% of assessors selected the target emotion and none of the foils were selected by more than 25% of the assessors [25].

The resulting sets of accepted videos for each of the three modalities (Upper Body, Face only, and Head only) and their intersections are shown in Figure 4.

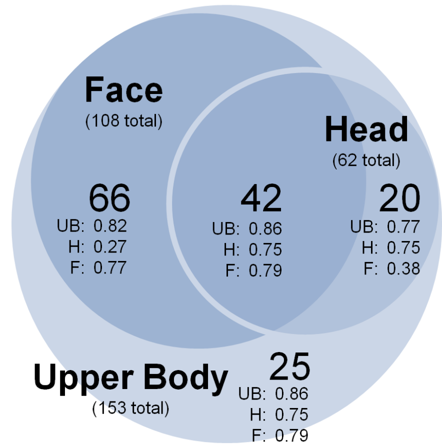


Fig. 4. The sets of accepted videos for each of the three modalities, along with the respective recognition rates for each modality for each set of videos (UB: Upper Body (head, face and shoulders), H: Head only, and F: Face only). There is a significant difference between the Head and Face recognition rates for the sets of videos that were recognizable by only the Head and by only the Face ($p<0.001$). More details are given in Table II and Figure 5.

We can see that certain videos are distinguishable through both the Head-only and the Face-only modalities (42 videos), some are distinguishable only through the Face (66 videos), some only through the Head (20 videos), while the rest were neither distinguishable by the Face alone nor the Head alone but required the full Upper Body to be identified (25 videos).

The recognition rates per modality for each set of accepted videos are shown in Table II and Figure 5. Statistical significance was calculated using a one-way ANOVA. There is no significant difference between the Head recognition rates and the Face recognition rates for the videos that neither modality could discriminate, nor for the videos where both modalities could discriminate.

Unsurprisingly, the Upper Body had a significantly higher recognition rate for videos that were neither recognizable by the Face nor the Head. Interestingly, the same was true for the videos that were recognizable by both the Head and the Face, providing evidence that the two modalities are able to reinforce each other to contribute positively to the overall recognition rate. Hence the emotional content in each modality is complementary rather than redundant.

The Head recognition rate for the set of Head-only accepted videos was significantly higher than the Face recognition rate for the same set of videos, and had no significant difference compared with the Upper Body recognition rate. This suggests that the emotional content in these videos is carried primarily by the head.

The Face recognition rate for the set of Face-only accepted videos was significantly higher than the Head recognition rate for the same set of videos. This suggests that the emotional

TABLE II
RECOGNITION RATES PER SET OF ACCEPTED VIDEOS
† = SIGNIFICANT DIFFERENCE COMPARED WITH HEAD
⋄ = SIGNIFICANT DIFFERENCE COMPARED WITH FACE
● = SIGNIFICANT DIFFERENCE COMPARED WITH HEAD

Recognition Set	**Number of Videos**	**Upper Body Recognition Rate**	**Head Recognition Rate**	**Face Recognition Rate**
Neither Head Nor Face	25	0.727 †⋄	0.316	0.338
Only Head	20	0.771 †	0.754 ●	0.384
Only Face	66	0.820 †⋄	0.270	0.766 ●
Both Head and Face	42	0.858 †⋄	0.753	0.794
Overall	**153**	**0.809** †⋄	**0.473**	**0.654** ●

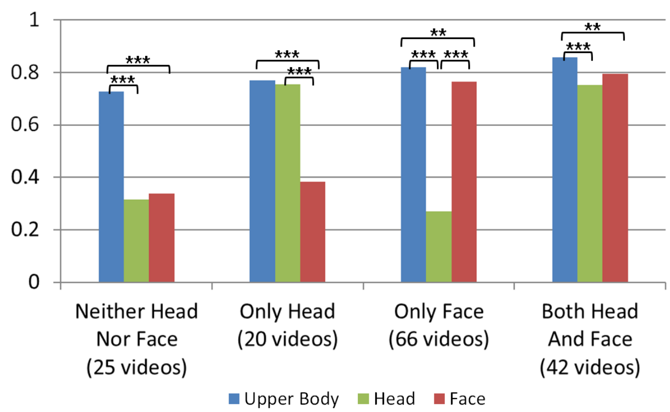


Fig. 5. Recognition rates for the four sets of accepted videos. The Upper Body recognition rates are significantly higher in nearly every set, suggesting that even when videos can be correctly recognized by both the Head alone and the Face alone there is still a reinforcing positive effect of seeing both modalities together. This, along with the significantly higher recognition rates for the 20 videos that were recognizable by the Head but not by the Face, suggests that the Head carries emotional information that is not redundant to the information in facial expressions. Statistically significant differences between modalities is denoted as ** for $p<0.01$ and *** for $p<0.001$.

content in these videos is present in the face, but is enhanced by including the full upper body (head and shoulders).

C. Recognition Rates Across Emotion Categories

Table III and Figure 6 present the recognition rates for each modality for each of the 18 emotion categories. We used a Related Samples Friedman Test with post-hoc t-tests with Bonferroni correction to determine the significant differences among modalities.

The Upper Body modality had significantly higher recognition rates than both the Head modality and the Face modality in the following emotion categories: *Afraid*, *Angry*, *Disappointed*, *Joking* and *Sad*. This suggests that these emotion categories require multiple modalities in order to be recognized correctly, supporting the work done by Hess et al. on anger and fear [23].

Conversely, *Bored*, *Happy*, *Sneaky* and *Surprised* can be recognized just as well from the Face alone as they can from the Upper Body (i.e. no significant difference in recognition rate), and *Interested* and *Frustrated* can be recognized just as well from the Head alone as they can from the Upper Body (i.e. no significant difference in recognition rate). This supports the link between interest and head pose, as suggested by Ekman [30] and Reeve [31].

Excited, *Hurt*, *Neutral*, and *Proud* can be detected just as well from the Face alone and from the Head alone as they can from the Upper Body (i.e. no significant difference in recognition rate between Face and Upper Body, and between Head and Upper Body). This supports previous work on raised heads for pride by Tracy and Robins [22].

The Face modality had significantly higher recognition rates than the Head modality for *Bored*, *Disgusted*, *Happy*, *Sneaky*, *Surprised* and *Worried*. The Head was not significantly better than the Face in any emotion categories, however the single *Ashamed* video in the corpus was far more recognizable by the Head than by the Face, supporting the previous work on bowed heads for shame and ‘inferiority emotions’ [15], [21].

These results support previous findings in psychology, and moreover, extend these findings to dynamic stimuli.

TABLE III
RECOGNITION RATES PER EMOTION CATEGORY
† = SIGNIFICANT DIFFERENCE COMPARED WITH HEAD
⋄ = SIGNIFICANT DIFFERENCE COMPARED WITH FACE
● = SIGNIFICANT DIFFERENCE COMPARED WITH HEAD

Emotion category	**Total Videos**	**Upper Body Recognition Rate**	**Head Recognition Rate**	**Face Recognition Rate**
Afraid	13	0.758 †⋄	0.359	0.476
Angry	7	0.791 †⋄	0.514	0.624
Ashamed	1	0.695	0.863	0.190
Bored	7	0.768 ⋄	0.318	0.724 ●
Disappointed	5	0.688 †⋄	0.327	0.361
Disgusted	11	0.830 †⋄	0.214	0.651 ●
Excited	9	0.823	0.646	0.737
Frustrated	8	0.808 †	0.651	0.468
Happy	12	0.832 ⋄	0.117	0.773 ●
Hurt	8	0.714	0.618	0.598
Interested	7	0.768 †	0.609	0.418
Joking	8	0.918 †⋄	0.576	0.810
Neutral	17	0.844	0.793	0.762
Proud	5	0.774	0.493	0.428
Sad	11	0.816 †⋄	0.535	0.602
Sneaky	4	0.797 ⋄	0.115	0.793 ●
Surprised	15	0.879 ⋄	0.467	0.804 ●
Worried	5	0.774 †⋄	0.312	0.889 ●
Overall	**153**	**0.809** †⋄	**0.473**	**0.654** ●

D. Nods and Shakes

We were also interested in investigating whether particular head motions carry more emotional information than others, namely whether head nods and shakes are the most influential head gestures in emotion perception.

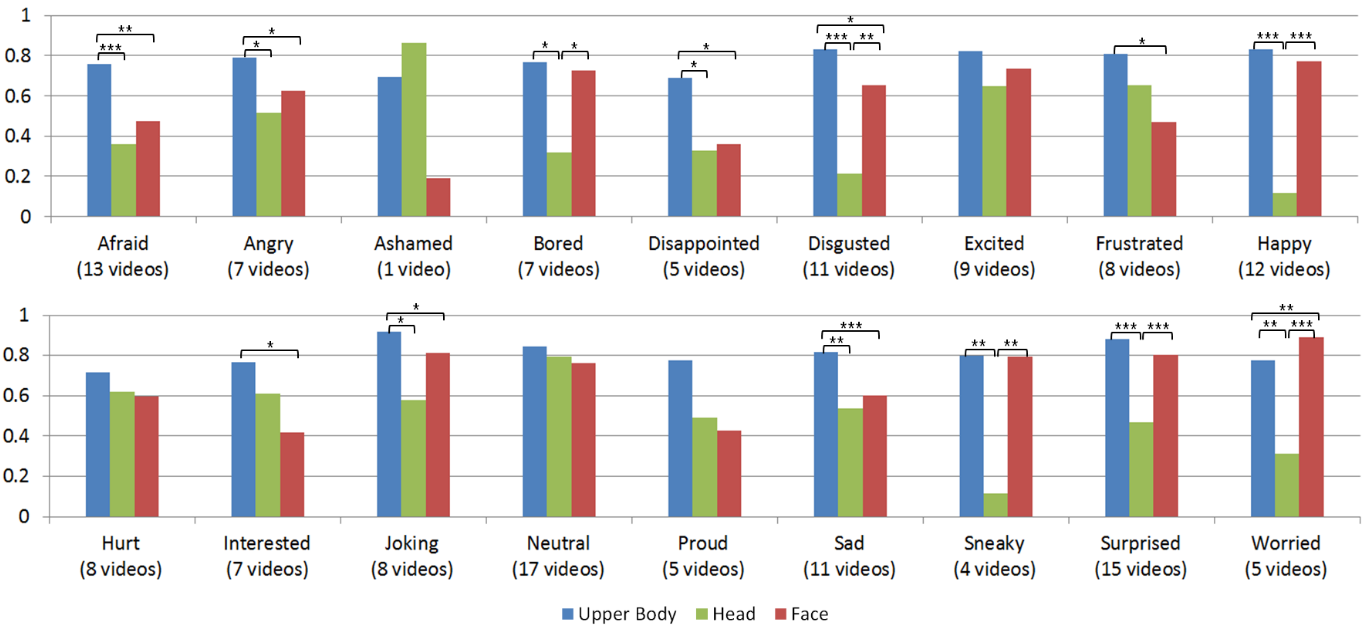


Fig. 6. Recognition rates for the 18 emotion categories. The Upper Body had significantly higher recognition rates than both the Head and the Face for *Afraid*, *Angry*, *Disappointed*, *Joking* and *Sad*, suggesting that these emotion categories require multiple modalities in order to be recognized correctly. Statistical significance between modalities is denoted as * for $p<0.05$, ** for $p<0.01$, and *** for $p<0.001$ (where p-values have been adjusted using Bonferroni correction).

The body of research linking emotion and head nods and shakes stretches back to Darwin [32]. More recently, nods and shakes have been frequently selected for automated detection in affective computing research [33], [34], while all other head gestures have largely been ignored.

1) Labelling of Nods and Shakes: The head-only version of each video was watched independently by two coders and labelled with one of the following categories: (1) contains nodding, (2) contains shaking, (3) contains nodding and shaking, (4) does not contain nodding or shaking. The two labellers agreed on 92% of the labels and discussed together the appropriate labelling for the other 8% of the videos.

The 109 videos that did not contain nodding or shaking were further classified by the same coders into two categories: (1) contains no head motion (i.e. continuously-held frontal head pose), or (2) contains head motion. The two labellers agreed on 88% of the labels and decided together the final labels for the other 12% of the videos. Videos with no head motion were primarily from the *Neutral* category (17 of the 27 videos).

Figure 7 depicts the presence of head gestures across each of the sets of accepted videos.

2) Recognition Rates for Nods and Shakes: The recognition rates per modality for each set of videos with particular head gestures are shown in Table IV and Figure 8. Statistical significance was calculated using a one-way ANOVA.

Only 17 of the 62 videos that were recognizable by the Head had nods and/or shakes; 18 had no motion, and the other 27 had other types of head motions (besides nods and shakes). Videos with no head motion were signficantly more recognizable for the Head when compared with videos with head motions other than nods and shakes.

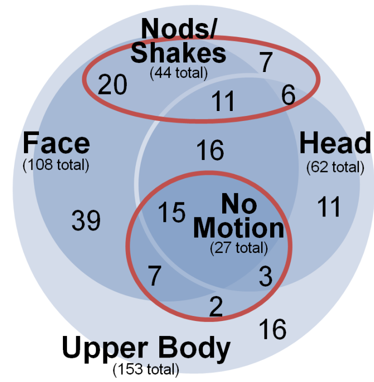


Fig. 7. Head gestures present across each of the sets of accepted videos. The videos containing head nods and/or shakes are indicated, as are the videos with no head motion. The videos that are not within these two sets therefore contain head motions that are neither nods nor shakes. Note the 27 videos that are recognizable by the Head modality even without nods and shakes.

In all three modalities, there is no significant difference in recognition rate for the videos which contain nods and/or shakes and the videos which contain other types of head motion. This suggests that there is as much emotional information in other types of head movements as there is in nodding and shaking of the head. It supports previous findings in psychology which link a variety of head motions with particular emotions (e.g. head lift to express pride [22] and sideways head tilt to express thinking [33]). It also suggests that a broader taxonomy of head gestures should be incorporated into the automated classifiers being developed in affective computing research.

TABLE IV
RECOGNITION RATES FOR VIDEOS WITH HEAD GESTURES
** DENOTES STATISTICALLY SIGNIFICANT DIFFERENCE BETWEEN 'NO HEAD MOTION' AND 'HEAD MOTION BUT NOT NOD/SHAKE' WHERE P<0.01.

	Nods and/or Shakes	**No Head Motion**	**Head Motion but Not Nod/Shake**
Number of Videos	44	27	82
Upper Body Recognition Rate	0.810	0.847	0.796
Head Recognition Rate	0.477	0.631 **	0.420
Face Recognition Rate	0.636	0.748	0.633

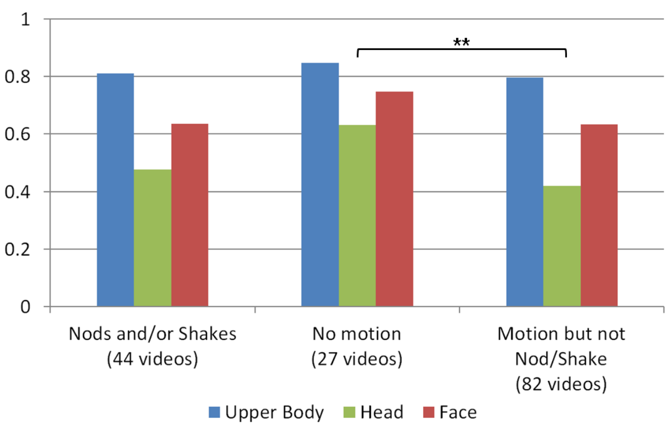


Fig. 8. Recognition rates per type of head gesture. In all three modalities, there is no significant difference in recognition rate for the videos which contain nods and/or shakes and the videos which contain other types of head motion. This suggests that there is as much emotional information in other types of head gestures as there is in nods and shakes. Statistically significant differences between modalities is denoted as ** for p<0.01.

E. Misrecognition Rates

Interestingly, the recognition rates of certain videos met the validation criteria (at least 50% agreement among raters, with none of the other responses having more than 25% agreement) for categories *other than* the target emotion.

The Face-only modality had only 4 videos which fit this misrecognition criteria, and no visible trends among them.

The Head-only modality had 11 videos that were misrecognized, including 7 of the 12 videos in the emotion category *Happy* which were mistaken for *Interested* (3 videos, 2 of which contained nodding) and *Bored* (4 videos, 1 of which had no motion). Since *Happy* is strongly identified from the face [35], this confusion suggests that the head is not enough to distinguish the valence of *Happy* from *Bored*, and to distinguish the arousal of *Happy* from *Interested*.

One Head-only video that was misrecognized had 93.3% recognition rate for Foil4 which is designed to be a strongly dissimilar emotion category [25]. In this case, a video from the category *Joking* was misrecognized as *Angry*. Perhaps the high arousal component in both emotions caused the confusion, similar to a previous finding that head nods (which are mostly associated with positive intent/emotions) might occur with feelings of rage as well [34]. However, confusion based on arousal may be limited to particular emotions or particular videos as no correlation was found across the overall set.

F. Limitations

We faced a set of challenges with our data collection. Firstly, it was sometimes difficult to decouple the head and the face in certain video frames during the pre-processing. For the head-only videos, sometimes slight facial movements could be inferred despite the facial masking, as the size and shape of the mask expanded or contracted with facial expressions. In a few cases, small parts of the shoulders or the forehead could be seen. Figure 9 shows three examples of video frames that were imperfectly processed.

Fig. 9. Sample frames from our videos showing some challenges in isolating modalities, such as unmasked forehead wrinkles, shoulders showing through, and blurring on warped faces due to extreme head motion and orientation.

Moreover, due to the imperfect video pre-processing and the relatively small set of videos available in the dataset, some emotion categories had a low number of videos. Analyzing a greater number of videos for each emotion category will yield more confidence in the results described above.

Lastly, since eye gaze was kept in the face-only videos, and eye gaze and head motion are strongly linked [23], [36], we cannot be sure of the effect of this partial decoupling.

V. CONCLUSION

We have presented the first quantitative study of the effect of decoupling facial expressions and head gestures in the perception of complex emotions from video. We compared our recognition results with prior emotion perception results of the same videos when the full upper body is present. We demonstrated that facial expressions and head motion contain complementary emotional information, and that head motion carries additional information that can generate significantly higher recognition rates in certain cases. We showed that there is as much emotional information in other types of head movements as there is in nodding and shaking of the head.

Future work should include the further decoupling of eye gaze from facial expressions, as well as the isolation and classification of specific head gestures to analyze their emotional content, possibly building a taxonomy of important head gestures. This work should also be extended to spontaneous affective displays.

ACKNOWLEDGEMENT

This work is supported by the Gates Cambridge Trust.

References

[1] N. Ambady and R. Rosenthal, "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis," *Psychological Bulletin*, vol. 111, no. 2, pp. 256–274, 1992.

[2] S. Duncan, "Some signals and rules for taking speaking turns in conversations." *Journal of personality and social psychology*, vol. 23, no. 2, p. 283, 1972.

[3] B. Parkinson, "Do facial movements express emotions or communicate motives?" *Personality and Social Psychology Review*, vol. 9, no. 4, pp. 278–311, 2005.

[4] P. Ekman, W. V. Friesen, and P. Ellsworth, *Emotion in the Human Face*, 2nd ed., P. Ekman, Ed. Cambridge University Press, 1982.

[5] F. De la Torre and J. F. Cohn, *Guide to Visual Analysis of Humans: Looking at People*. Springer, 2011, ch. Facial Expression Analysis.

[6] J. F. Cohn and F. De la Torre, "Automated face analysis for affective computing," *The Oxford Handbook of Affective Computing*, p. 131, 2014.

[7] J. N. Bassili, "Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face." *Journal of personality and social psychology*, vol. 37, no. 11, p. 2049, 1979.

[8] Z. Hammal and J. F. Cohn, "Intra-and interpersonal functions of head motion in emotion communication," in *Workshop on Roadmapping the Future of Multimodal Interaction Research including Business Opportunities and Challenges*. ACM, 2014, pp. 19–22.

[9] R. el Kaliouby, P. Robinson, and S. Keates, "Temporal context and the recognition of emotion from facial expression," in *HCI International Conference*, 2003.

[10] M. F. Valstar, "Timing is everything: A spatio-temporal approach to the analysis of facial actions," Ph.D. dissertation, 2008.

[11] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.

[12] G. A. Ramirez, T. Baltrušaitis, and L.-P. Morency, "Modeling latent discriminative dynamic of multi-dimensional affective signals," in *1st International Audio/Visual Emotion Challenge and Workshop in conjunction with Affective Computing and Intelligent Interaction*, 2011.

[13] H. Gunes and M. Pantic, "Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners," in *Intelligent virtual agents*. Springer, 2010, pp. 371–377.

[14] P. Ekman and W. V. Friesen, "Head and body cues in the judgment of emotion: A reformulation," *Perceptual and motor skills*, vol. 24, no. 3, pp. 711–724, 1967.

[15] J. L. Tracy and D. Matsumoto, "The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays," *Proceedings of the National Academy of Sciences*, vol. 105, no. 33, pp. 11 655–11 660, 2008.

[16] D. Keltner, "Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion," *Cognition & Emotion*, vol. 10, no. 2, pp. 155–172, 1996.

[17] J. A. Harrigan and D. M. O'Connell, "How do you look when feeling anxious? Facial displays of anxiety," *Personality and Individual Differences*, vol. 21, no. 2, pp. 205–212, 1996.

[18] A. N. Meltzoff and M. K. Moore, "Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms." *Developmental psychology*, vol. 25, no. 6, p. 954, 1989.

[19] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, "Automated analysis of body movement in emotionally expressive piano performances," 2008.

[20] I. Poggi, F. D'Errico, and L. Vincze, "Types of nods. The polysemy of a social signal." in *LREC*, 2010.

[21] A. Mignault and A. Chaudhuri, "The many faces of a neutral face: Head tilt and perception of dominance and emotion," *Journal of Nonverbal Behavior*, vol. 27, no. 2, pp. 111–132, 2003.

[22] J. L. Tracy and R. W. Robins, "The prototypical pride expression: Development of a nonverbal behavior coding system." *Emotion*, vol. 7, no. 4, p. 789, 2007.

[23] U. Hess, R. B. Adams Jr, and R. E. Kleck, "Looking at you or looking elsewhere: The influence of head orientation on the signal value of emotional facial expressions," *Motivation and Emotion*, vol. 31, no. 2, pp. 137–144, 2007.

[24] D. Neth and A. M. Martinez, "Emotion perception in emotionless face images suggests a norm-based representation." *Journal of vision*, 2009.

[25] H. O'Reilly, D. Pigat, S. Fridenson, S. Berggren, S. Tal, O. Golan, S. Bölte, S. Baron-Cohen, and D. Lundqvist, "The EU-Emotion Stimulus Set: A Validation Study," *Under review*, 2015.

[26] M. G. Frank and J. Stennett, "The forced-choice paradigm and the perception of facial expressions of emotion." *Journal of personality and social psychology*, vol. 80, no. 1, p. 75, 2001.

[27] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Continuous conditional neural fields for structured regression," in *European Conference on Computer Vision*. Springer, 2014, pp. 593–608.

[28] K. Krippendorff, "Agreement and information in the reliability of coding," *Communication Methods and Measures*, vol. 5, no. 2, pp. 93–112, 2011.

[29] S. Afzal and P. Robinson, "Natural affect data: Collection and annotation," in *New Perspectives on Affect and Learning Technologies*. Springer, 2011, pp. 55–70.

[30] P. Ekman and H. Oster, "Facial expressions of emotion," *Annual review of psychology*, vol. 30, no. 1, pp. 527–554, 1979.

[31] J. M. Reeve, "The face of interest," *Motivation and Emotion*, vol. 17, no. 4, pp. 353–375, 1993.

[32] Darwin, Charles (1872), *The expression of the emotions in man and animals*. Oxford University Press, 2002.

[33] R. El Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," in *Real-time vision for human-computer interaction*. Springer, 2005, pp. 181–200.

[34] A. Kapoor and R. W. Picard, "A real-time head nod and shake detector," in *Proceedings of the Workshop on Perceptive User Interfaces*. ACM, 2001, pp. 1–5.

[35] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti *et al.*, "Universals and cultural differences in the judgments of facial expressions of emotion." *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.

[36] J. F. Cohn, L. I. Reed, T. Moriyama, J. Xiao, K. Schmidt, and Z. Ambadar, "Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles," in *6th IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2004, pp. 129–135.