Measuring Voter's Candidate Preference Based on Affective Responses to Election Debates

Daniel McDuff[†‡], Rana El Kaliouby[†], Evan Kodra[†] and Rosalind Picard[†‡]
† Affectiva, Waltham, MA 02452, USA
‡ MIT Media Lab, Cambridge, MA 02139, USA
Email: {djmcduff, picard}@media.mit.edu, {kaliouby, evan.kodra}@affectiva.com

***Abstract*—In this paper we present the first analysis of facial responses to electoral debates measured automatically over the Internet. We show that significantly different responses can be detected from viewers with different political preferences and that similar expressions at significant moments can have very different meanings depending on the actions that appear subsequently. We used an Internet based framework to collect 611 naturalistic and spontaneous facial responses to five video clips from the 3rd presidential debate during the 2012 American presidential election campaign. Using this framework we were able to collect over 60% of these video responses (374 videos) within one day of the live debate and over 80% within three days. No participants were compensated for taking the survey. We present and evaluate a method for predicting independent voter preference based on automatically measured facial responses and self-reported preferences from the viewers. We predict voter preference with an average accuracy of over 73% (AUC 0.779).**

I. Introduction

Political debates cover emotive issues that impact people's lives. The policies the candidates present and the way in which they present them can have a significant bearing on their public perception and potentially on the outcome of the election. In the 2012 US Presidential election campaign there were three live televised debates between the presidential candidates Barak Obama and Mitt Romney. These each lasted one and a half hours and covered foreign and domestic policy. In this paper we present analysis of naturalistic and spontaneous responses to video segments of the third presidential debate. We show that significantly different responses to the candidates are measurable using automated facial expression analysis and that these differences can predict self-report candidate preference. We also identify moments within the clips at which initially similar expressions are seen but the temporal evolution of the expressions leads to very different associations.

Advertising is a huge component of political campaigns. In their 2012 US Presidential election campaigns the democratic and republican parties spent almost $1,000,000,000 on TV advertising[1]. A number of studies have focused on responses to political advertisements in order to measure their effectiveness and evaluate the emotional responses of viewers. These studies have generally focused on self-reported emotions. Political debates cover many of the same campaign themes as political advertising and the techniques presented in this paper would generalize to the evaluation of political ads. Kraus [1] provides detailed history on televised presidential debates which are viewed as a key element in modern campaigns.

[1]http://www.washingtonpost.com

Fig. 1. Top) Images from the election debate between President Barack Obama and Governor Romney. Bottom) Images taken subset of facial responses data collected with permission to share publicly.

Brader [2] found that political campaigning - in particular TV advertising - achieves its goal in part by appealing to the emotions of the viewers and that different emotional states led to different self-report responses. In particular, whether an ad appeals to fear or enthusiasm can have a considerable effect on its persuasive power. The success of political advertising is typically measured by polling audiences. This may be performed via telephone interviews or focus groups. Luntz [3] highlights the power in audience measurement. However, he also identifies that focus groups, in which people gather in a single room and report their feelings via a button- or dial-operated computer, can be the least financially profitable tool in political polling. Focus groups present many challenges, reporting emotions using a dial or slider can detract from the experience of interest and participants are typically limited to those within a small geographic area. The Internet framework we present is vastly less expensive than this method of polling and has several other benefits. Self-report methods are susceptible to cognitive biases and can distract attention from the media being evaluated. Measuring affective responses from the face allows for rich multi-dimensional temporal data to be collected without an additional task or specialized hardware. Only a standard webcam and Internet are required by viewers and the time commitment is reduced as the task can be

performed from home. In addition, none of the viewers in this study were compensated for their time.

The human face is a powerful channel for communicating valence as well as a wide gamut of emotion states. The Facial Action Coding System (FACS) [4], [5], [6] is a comprehensive catalogue of unique action units (AU) that correspond to each independent motion of the face. FACS enables the measurement and scoring of facial activity in an objective, reliable and quantitative way, and is often used to discriminate between subtle differences in facial motion [7]. Facial behavior has been used to measure the effectiveness of media content, typically in the form of short advertising video clips [8], [9].

The main contributions of this work are:

1) To present a dataset of naturalistic and spontaneous facial responses, collected over the Internet, to segments from a real election debate.
2) To identify salient segments of the content based on facial emotion responses and show that expressions contain detailed temporal information. For example: symmetric AU12 (smirks) followed by smiles signals a different state than those not followed by a smile.
3) To design and evaluate a method for candidate preference prediction based on temporal facial responses.

The remainder of the paper will discuss the data collection, experiment, insights from the facial responses, modeling and prediction results. This paper is also supported by supplementary material. The data and results presented here can be understood in much greater detail by viewing and interacting with the data live. This includes: the stimuli videos in full, transcripts of the clips and an interactive dashboard for exploring the data. The material can be found at: http://blue-labs.affectiva.com/afdebate/.

II. Related Work

Previous work has shown that it is possible to efficiently collect a large number of facial responses to media online [10]. Further to this, it is possible to accurately predict viewer liking and desire to view again of video based on automatically smile responses measured over the Internet [11]. In addition, Kodra et al. [6] showed that continuous measurement of facial behavior correlated highly with self-reported(dial) measures of media preference.

Automated facial expressions analysis is a large field of research that combines understanding in psychology with computer vision and machine learning algorithms for detecting facial expressions. The majority of automated facial analysis systems detect facial action units or discrete emotional states (typically six states: amusement, fear, anger, disgust, surprise and sadness). De La Torre and Cohn [12] present a summary of state of the art approaches to automated action unit detection. A number of approaches [13], [10] use the assumption that the probability estimate from the classifiers (e.g. distance from the SVM hyper-plane) correlate with the intensity of the action unit or expression.

III. Data Collection

Using a web-based framework similar to that described in [10] we collected facial responses to five video clips from the 3rd presidential debate during the 2012 presidential election campaign in the USA. The debate focused on foreign policy. A short description and lengths of the five clips chosen are shown in Table I. These were chosen as they covered significant issues and contain significantly different views from each of the candidates. In total 237 people opted-in and completed all or part of the experiment (they were not obliged to watch all clips) and 917 video responses were recorded. For the automatic analysis here videos for which it was not possible to identify and track a face in at least 20% of frames were disregarded; this left 611 videos (67%). From this point forward we only consider those videos that were trackable for greater than 20% of the frames.

TABLE I. Synopses and lengths of the five debate clips that viewers watched during the survey. The number of trackable video responses collected are also shown.

Clip	Duration	Description	Responses
1	55s	**Criticism of U.S. Navy**: President Obama responds to Governor Romney's criticism that the US Navy has fewer ships than it has since 1916.	154
2	53s	**Tour of the Middle East**: Governor Romney comments on President Obama's tour of the Middle East early in his presidency.	119
3	60s	**Bin Laden Assassination**: President Obama speaks about the personal impact of killing Osama bin Laden.	120
4	43s	**Trade War with China**: Governor Romney responds to a question about whether he would start a trade war with China.	111
5	68s	**Fate of Detroit**: The candidates spar over the fate of Detroit and the auto industry.	107

The website was launched the day following the debate (23rd October 2012) and was promoted on the front page of the New Scientist website (http://www.newscientist.com/), most participants reached it via this link or via a link on a social networking site. Figure 2 shows the number of facial responses to the clips that were collected on the launch day as well as subsequent days. Using this framework we were able to collect over 60% of the video responses (374 videos from 98 people) within one day of the live debate and over 80% (501 from 135 people) within three days. This represents a very efficient method of measuring responses - necessary when responses to the material may be time sensitive as with topical debates. No participants were compensated for taking the survey. The participants were also given the opportunity to share their data for research purposes, 95 people (40%) chose to do so. Anyone with at compatible browser and webcam could take part and as all the data was collected over the Internet we were able to obtain responses from people from a wide range of locations and demographic profiles. People in over 19 countries took part with 62% in the US. The demographics are summarized in Table II. Of the viewers (37%) declared a Democratic party affiliation. As the survey was open to the public the participants were not limited to be in the US or to be eligible to vote in the election. For the modeling of voter preferences we disregard those participants who said they were ineligible to vote.

All videos were recorded with a resolution of 320x240 and a frame rate of 15 fps. Participants were aware from the opt-

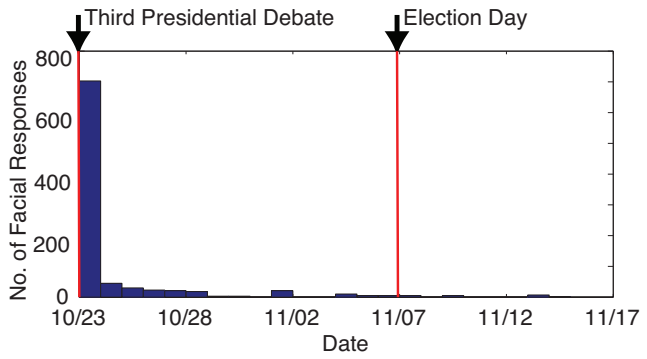

Fig. 2. The number of participants who took part in the survey and watched at least one of the debate clips.

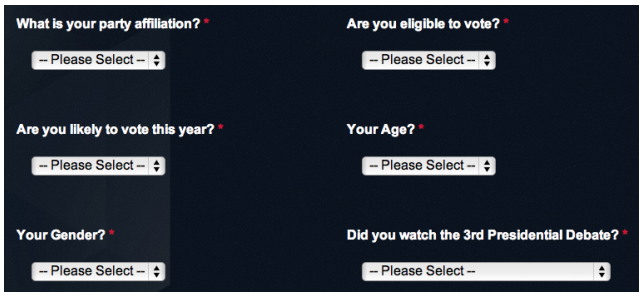

Fig. 3. Questions asked before the participants watched the debate clips. Responses to all questions were required.

in menu that their camera would be turned on and this may have had an impact on their facial response. However, people responded naturally in a vast majority of cases.

In order to collect data for this work we created a web-based survey which required people to answer multiple choice questions related to their party affiliation, candidate preferences, familiarity with the debate and demographic profile. Following this they watched five clips from the debates and their facial responses were recorded using the framework described above. Figure 3 shows a screenshot of the questions asked before the viewers watched the clips. Following each clip viewers were asked the following question: "After watching this clip, which candidate do you prefer?". Viewers were required to pick either "President Barack Obama" or "Governor Mitt Romney."

TABLE II. DEMOGRAPHIC PROFILE OF THE PARTICIPANTS IN THE 611 FACIAL RESPONSES VIDEOS COLLECTED.

Ages (years)	18-25: 168 (27%), 26-35: 165 (27%), 36-45: 138 (23%), 46-55: 79 (13%), 56-65: 39 (6%), 65+: 22 (4%)
Gender	Male: 407 (67%), Female: 204 (33%)
Location	United States: 378 (62%), Outside US: 233 (38%) Australia (11), Bulgaria (1), Canada (10), China (6), Czech Rep. (5), Denmark (6), Egypt (11), Estonia (4), Germany (13), Honduras (11), Italy (8), Rep. of Korea (1), Netherlands (13), Norway (5), Sweden (3), Switzerland (5), UK (113).
Eligibility to Vote	Yes: 415 (68%), No: 196 (32%)
Party Affiliation	Democratic: 226 (37%), Republican: 51 (8%), Independent: 150 (25%), None: 184 (30%)

IV. AUTOMATED FACIAL EXPRESSION ANALYSIS

We developed automated algorithms for smirks, smiles and valence detection. In order to detect the expressions the Nevenvision facial feature tracker[2] is first used to automatically detect the face and tracks 22 facial feature points on each frame of the video. For each AU, a region of interest (ROI) around the appropriate part of the face is located using the landmark points. The image is cropped to the ROI and image features extracted. The exact details of the methods used for each of the algorithms are described below. The classifiers were trained and tested on webcam images similar in quality to the webcam videos analyzed in this study. The data was divided into three datasets, 50% used for training, 17% used for validation (multiple potential classifiers are evaluated and the best selected) and 33% used for testing. To ensure person-independent experiments, frames from a particular facial video were used exclusively to train or to test the system.

Smirks: We define smirks as presence of an asymmetric AU12 (Zygomatic major). For training and testing the smirk detector 5,100 labeled example images were used. The area under the ROC curve during testing was 0.88. More details about the smirk detector used can be found in [14].

Smiles: The smile classifier uses an ROI around the mouth and a Support Vector Machine (SVM) with RBF kernel. For training and testing the smile detector 15,700 labeled example images were used. The area under the ROC curve during testing was 0.97.

Valence: Valence (V), defined as a measure for the overall positivity of a person's facial state, is calculated using HOG features extracted from the whole face, which are then input to a Support Vector Regression (SVR) with an RBF kernel. The output ranges from -1 to 1, where -1 indicates negative valence and 1 positive valence. V is trained using a three class system. Positive valence (+1) is defined as the presence of a smile and negative valence (-1) defined as non-positive valence images with the presence of AU4 or AU9. All other images are defined as valence neutral (0). 65,000 labeled example images were used to train and test the valence detector. The area under the ROC curve during testing was 0.90.

Similar facial coding algorithms have been used in other applications. Further details on training, testing and validation of the facial coding classifiers can be found in [6].

V. INSIGHTS FROM FACIAL RESPONSES

In the first section of this analysis we look at the differences in aggregate responses between those that reported a preference for Obama or Romney following each clip. Figure 5 shows the mean valence measured for those that reported preference for Obama (blue) and Romney (red) after watching the clips. Immediately the facial responses tell us detailed information about the parts of the clips that had greatest emotional response. Both candidates appear to have scored points during the debates, in particular, Obama during clip one and Romney during clip five. The greatest difference in measured facial activity between the two groups (those who preferred Obama versus those who preferred Romney) occurred when Obama made a joke about why the Navy was

[2]Licensed from Google, Inc.

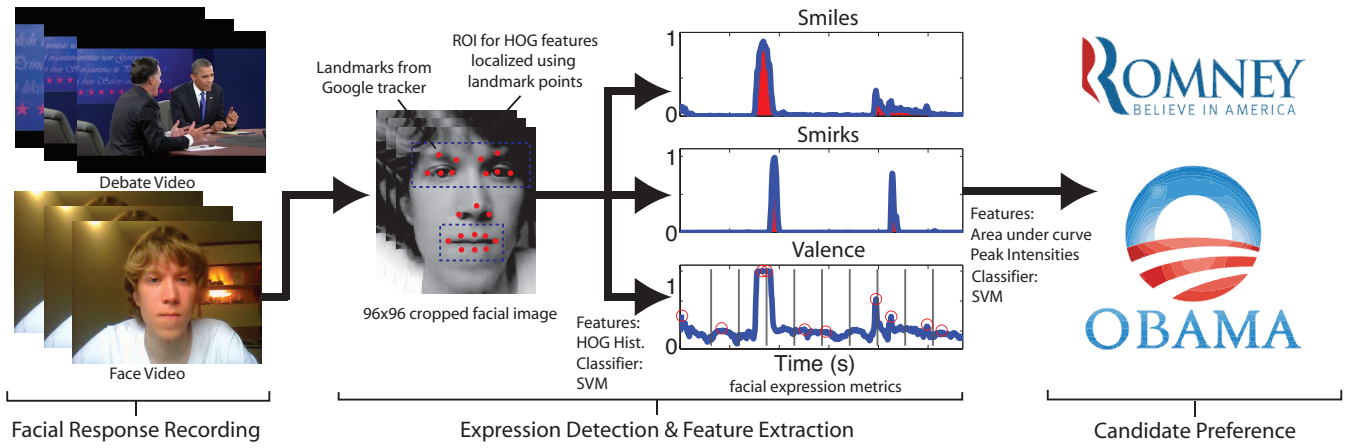

Fig. 4. A system diagram of our facial expressions and candidate recognition system. The facial videos are recorded whilst the viewer watches the debate. Each frame is extracted and smile and smirk probabilty and valence calculated. Features are extracted from the resulting time series.

TABLE III. EXCERPTS FROM THE DEBATE TRANSCRIPT WHICH CORRESPOND TO THE MARKED POINTS ON FIGURE 5. THE COLOR OF THE DOTS REFERS TO THE SPEAKER: BLUE=OBAMA, RED=ROMNEY.

Ref	Quote
a	The nature of our military's changed. We have these things called aircraft carriers where planes land on them.
b	We have these ships that go underwater, nuclear submarines
c	One of the challenges we've had with Iran is that [people] felt that the administration was not as strong as it needed to be. I think they saw weakness where they had expected to find American strength.
d	Then when there were dissidents in the streets of Tehran, the Green Revolution, holding signs saying, is America with us, the president was silent. I think they noticed that as well.
e	Obama Interupts: "Governor Romney, that's not what you said."
f	Under no circumstances would I do anything other than to help this industry get on its feet. And the idea that has been suggested that I would liquidate the industry, of course not. Of course not.

not investing in more ships - quotes a and b in Table III. Another highlight occurs when the candidates discuss the decline of the car industry in detroit. Obama interrupts Romney (e) but it seems that Romney's response is sufficient.

The next section of the analysis that we performed was to look at the moments during the debate clips that cause the greatest amount of facial behavior - as detected by our expression detection algorithms described above. Figure 6 shows examples in the responses to clip 1 and clip 4 during which the viewers smirked. However, in the two clips the relationship between the smirks and smiles was very different. During the first clip the smirks were followed in most cases by a symmetric AU12 or smile. However, during clip 4 the smirks were followed by more varied responses including AU1+2, AU25 and AU4 - generally suggesting a more negative valence than smiles [15].

VI. PREDICTING VOTER PREFERENCE

We present a method for predicting voter's candidate preference from responses to the debate clips. We perform a person independent test and use an SVM for classification. The modeling was performed by removing data from 10 participants from the data for a particular debate clip and then performing a leave-out-one validation on the remaining data. The classifier was then trained using the validated parameters and testing on the withheld data. As the class sizes were unbalanced, we over-sampled the testing set in each case to reflect an equal distribution of Obama and Romney labels. The sampling was performed by selecting an equal number of samples from each class in the testing set.

A particularly interesting subset of the voting population are those that identify themselves as independent. Campaigns for the democratic and republican candidates often focus large amounts of effort on winning the vote of these people. We show the results for this population in particular.

A. Features

In order to train and test the preference classifier we extracted features from the raw metrics. The features were calculated from the valence metric as this is a combination of both positive and negative expressions. We divided each responses into 10 evenly spaced temporal bins and took the maximum valence peak within each bin, a similar method to that used by McDuff et al. [11] to predict liking preference. As additional features we took the area under the smirk track and the smile track. This gave a final feature vector of length 12. Figure 4 (c) shows an example of the features extracted. The features extracted were normalized using a z-transform (to result in zero mean and unit standard deviation), this was performed across the data in order to improve generalizability as the clips were of different lengths.

B. Model

Support Vector Machines (SVM) are a static approach to classification and therefore do not explicitly model temporal dynamics. However, as described above the features extracted from the data captured the dynamics. A Radial Basis Function (RBF) kernel was used. The model parameters were validated using a leave-one-out procedure on the training set. During validation the penalty parameter, C, and the RBF kernel parameter, γ, were each varied from 10^k with k=–2, 1,..., 2.

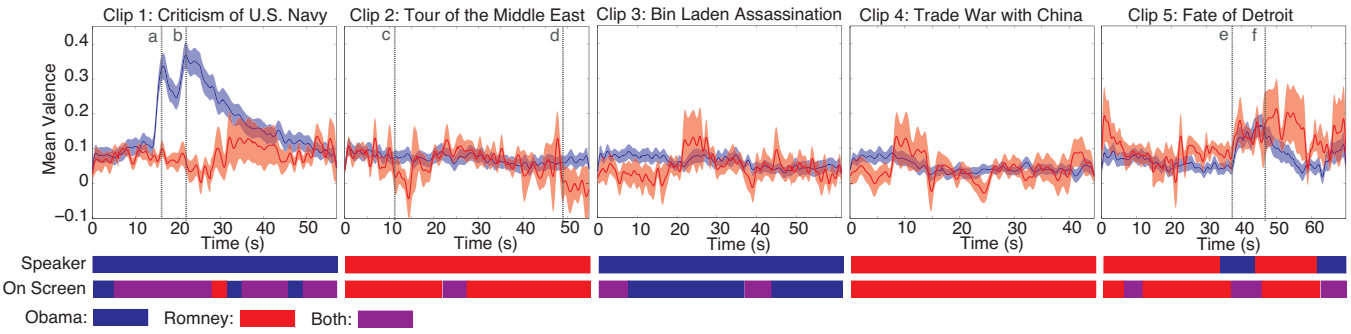

Fig. 5. Mean valence during the clip for those that reported a preference for Obama (blue) and Romney (red) after watching the clip. The shaded area represented the standard error range. Below the plots we show which candidate was speaking and which (or both) was on screen during the clip. The letters and dotted lines correspond to significant parts of the clips - transcripts of these parts of the clips can be found in Table III.

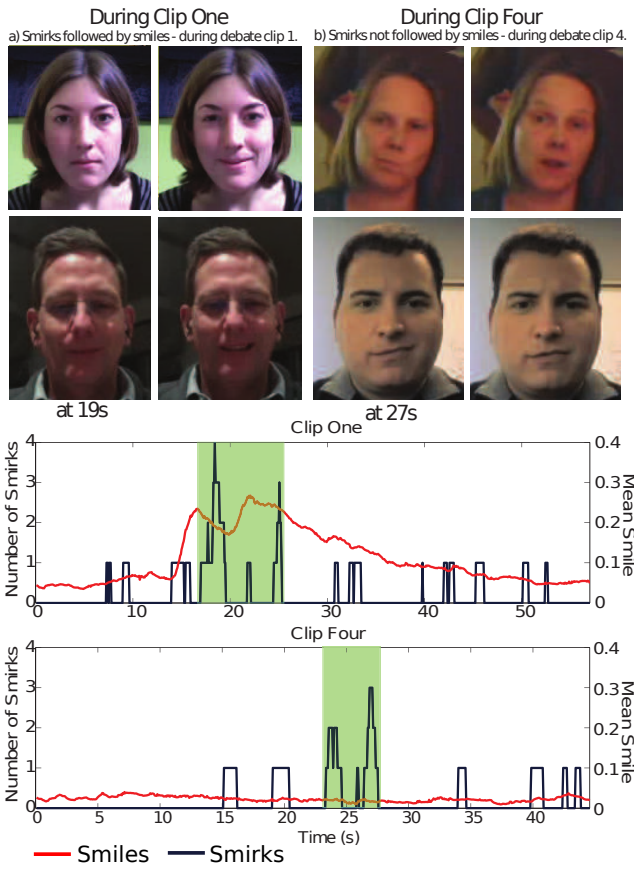

Fig. 6. Top) Examples of smirks during two parts of the debate clips. Top left) Smirks that occurred during Clip1which were followed by smiles/positive valence. Top right) Smirks that were not followed by smiles/positive valence. Bottom) Plots of the aggregate smiles and number of smirks for debate Clip 1, and debate Clip 4. Regions in which a greater number of smirks occurred are highlighted in green.

The SVM's were implemented using libSVM [16]. The median parameters selected during the model validation were $\gamma = 0.1$ and $C = 10$.

VII. Results and Discussion

A. Voter Preferences

Table IV (top) shows the confusion matrix for the prediction of candidate preferences from the facial valence information. The accuracy of the model was 73.8%. The model represents a significant improvement, over a naive model, for which the accuracy would be 50%. The Precision-recall and ROC curves (green lines) are shown in Figure 7, the area under the ROC curve was 0.818.

TABLE IV. Top) Confusion matrix for prediction of voter preference across all the eligible voters. Bottom) Confusion matrix for prediction of voter preference across the eligible voters with no or an independent party affiliation. Threshold determined as the case closest to the ROC (0,1) point.

		Actual outcome: Obama	Actual outcome: Romney	Total
Predicted value	Obama′	39.9%	16.1%	56%′
	Romney′	10.1%	33.9%	44%′
	Total	50%	50%	

		Actual outcome: Obama	Actual outcome: Romney	Total
Predicted value	Obama′	40.5%	15.5%	56%′
	Romney′	9.5%	34.5%	44%′
	Total	50%	50%	

B. Independent Voter Preferences

Table IV (bottom) shows the confusion matrix for the prediction of candidate preferences from the facial valence information of self-reported independent voters. The Precision-recall and ROC curves (blue lines) are shown in Figure 7, the area under the ROC curve was 0.733. The prediction accuracy is still strong (75%) and the area under the precision-recall and ROC curves are slightly greater 0.841 and 0.801 respectively. This is encouraging as the preferences of independent voters are perhaps of more interest than those with a declared democrat or republican affliation. In particular, a higher percentage of those that preferred Obama were classified as preferring Romney.

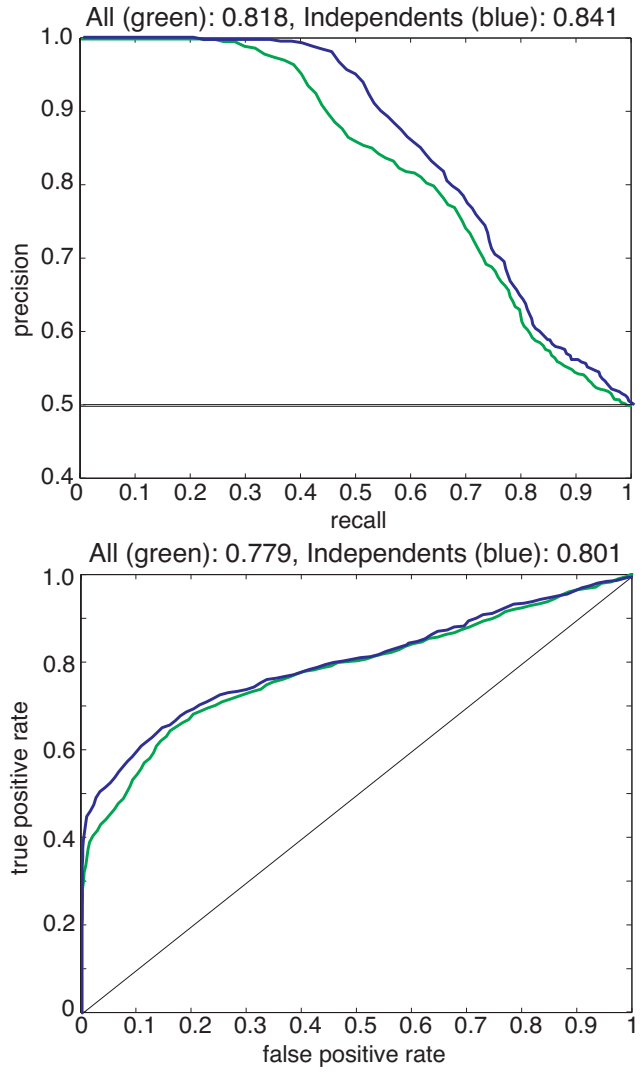

Fig. 7. Precision-recall (top) and ROC curves (bottom) for the voter preference prediction task. Green) Results for all eligible voters. Blue) Results for all eligible voters with no or an independent party affiliation.

VIII. Conclusions and Future Work

We used an Internet based framework to collect 611 naturalistic and spontaneous facial responses to five video clips from the 3rd presidential debate during the 2012 American presidential election campaign. Using the Internet as a channel allows for data to be collected very efficient and quickly, which is especially important when considering live topical events such as election debates. We were able to collect over 60% of these video responses (374 videos) within 1 day of the live debate and over 80% (501) within three days.

We show that different responses can be detected from viewers with different political preferences and that similar expressions at significant moments appear to have very different meanings. In particular, we automatically identify asymmetric AU12 (smirks) at particular moments, the interpretation of which is highly dependent on the subsequent expressions. A model for predicting candidate preference based on automatically measured facial responses is presented and we show that its accuracy is significantly greater than a naive model. The ROC AUC for the model is 0.779 for voters with a democrat or republican affiliation and 0.801 for just the independent voters with neither a democrat or republican affliation.

This work presents much potential for the application of emotional measurement to political ads, debates and other media. Other work has shown the power of facial responses to advertising in predicting effectiveness [8], [11]. This could be extended to specifically political advertising and the nuances associated with that domain. Future work will also address the automated mining of salient sequences of actions.

IX. Acknowledgements

We would like to thank Melissa Burke, Elina Kanan, May Amr and Andy Dreisch for contributions to this work. We would also like to thank the viewers who took part in the study. Lisa Grossman and the New Scientist generously helped promote the online data collection.

References

[1] S. Kraus, *Televised presidential debates and public policy*. Lawrence Erlbaum, 1999.

[2] T. Brader, "Striking a responsive chord: How political ads motivate and persuade voters by appealing to emotions," *American Journal of Political Science*, vol. 49, no. 2, pp. 388–405, 2005.

[3] F. Luntz, "Focus group research in american politics," *The polling report*, vol. 10, no. 10, p. 1, 1994.

[4] P. Ekman and W. Friesen, "Facial action coding system," 1977.

[5] J. Cohn, Z. Ambadar, and P. Ekman, *Observer-based measurement of facial expression with the Facial Action Coding System*. Oxford: NY, 2005.

[6] E. Kodra, T. Senechal, D. McDuff, and R. Kaliouby, "From dials to facial coding: Automated detection of spontaneous facial expressions for media research," in *Automatic Face & Gesture Recognition and Workshops (FG 2013), 2013 IEEE International Conference on*. IEEE, 2013.

[7] C. Hjortsjö, *Man's face and mimic language*. Studen litteratur, 1969.

[8] R. Hazlett and S. Hazlett, "Emotional response to television commercials: Facial emg vs. self-report," *Journal of Advertising Research*, vol. 39, pp. 7–24, 1999.

[9] T. Teixeira, M. Wedel, and R. Pieters, "Emotion-induced engagement in internet video ads," *Journal of Marketing Research*, 2010.

[10] D. McDuff, R. El Kaliouby, and R. Picard, "Crowdsourcing facial responses to online videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 4, pp. 456–468, 2012.

[11] ——, "Predicting online media effectiveness based on smile responses gathered over the internet," in *Automatic Face and Gesture Recognition, 2013. Proceedings. Tenth IEEE International Conference on*, 2013.

[12] F. Torre and J. Cohn, "Facial expression analysis," *Visual Analysis of Humans*, pp. 377–409, 2011.

[13] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (cert)," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 298–305.

[14] T. Senechal, J. Turcot, and R. El Kaliouby, "Smile or smirk? automatic detection of spontaneous asymmetric smiles to understand viewer experience," in *Automatic Face and Gesture Recognition, 2013. Proceedings. Tenth IEEE International Conference on*, 2013.

[15] D. McDuff, R. El Kaliouby, K. Kassam, and R. Picard, "Affect valence inference from facial action unit spectrograms," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, pp. 17–24.

[16] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.