User Behavior Analysis for Online Advertising by Facial Expression Responses

Songfan Yang* Le An†

Abstract

The surge of multimedia contents being made available online has attracted more and more users. As a new platform, more advertisements have been presented online due to the increased user population and reduced advertising cost. In addition, a large portion of revenue for major IT companies such as Google and Facebook is generated by online advertisements. Hence, for both advertiser and advertisement hosts, effective advertisement targeting and delivering is of great interest and importance. In this paper, we quantify users' advertisement viewing experience based on their facial expression responses and propose a metric termed moment-to-moment zapping probability (MMZP) for predicting users' *zapping*, i.e., skipping behavior. Insights are discovered by analyzing MMZP and such knowledge may help improve the advertising effectiveness,e.g., presenting the most suitable advertisements to the user with specific attributes such as gender and ethnicity.

1 Introduction

With the proliferation of online multimedia contents as well as the rapid growth of Internet users, broadcasting and advertising business are shifting their attention from traditional media such as TV to the Internet. As an example, the revenue from Internet advertising is expected to boost over 60% to nearly 200 billion USD from 2013 to 2018. Due to the increased audience population and reduced cost of online advertising, advertisers are more inclined to publish advertisements (ads) online. On the other hand, most popular Internet sites (e.g., Google, Facebook) depend on advertising as their major revenue source. As a result, effective advertising are important to both advertisers and advertisement hosts.

To evaluate the quality of advertising, user behavior is usually studied. Such analysis can be conducted using users' self-reported feedback. However, this approach suffers from cognitive bias and cannot provide users' dynamic feedback to the ad provider [17]. Another way of evaluating ad quality is to study users' *zapping* behavior. Zapping is an important metric which is

Figure 1: For online advertisement, users often have option to *zap*, i.e., skip the advertisement.

defined as the action that a user stops watching an ad. Zapping implies diminished attention of the user who would less likely become a potential consumer [6]. In traditional media such as TV, zapping occurs when the user switches channel. For online advertising, the user normally has the option to zap (i.e., skip) an ad, as shown by an example in Fig. 1. Thus, accurately predicting and preventing zapping can lead to more effective advertising campaign.

As suggested in [1], user behavior is affected by perceptual considerations such as personal feelings; extracting such information is a challenging task. In this paper we analyze the user behavior during ad watching using their facial expression responses. As a rich source of implicit communication, facial expression reflects one's spontaneous feeling and can be measured non-intrusively [5]. Automatic facial expression recognition directly benefit many practical applications such as human behavior analysis, human-human interaction and human-computer interaction. Particularly, facial expression based analysis is a preferred way to evaluate the effectiveness of an ad during its course of playing, and such analysis can be performed dynamically while being non-intrusive [19]. With accurate facial expression analysis, marketing and advertising researchers can better understand a user's emotional state and behavior and consequently strategies can be planned to improve the effectiveness of advertising or even design interactive ads to enhance the users' experience. Some recent works have introduced facial expression analysis for the prediction of zapping behavior [27, 25].

*College of Electronics and Information Engineering, Sichuan University, Chengdu 610064, China. (e-mail: syang@scu.edu.cn)

†University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. (email:lan004@unc.edu)

In this paper, to study the users' behavior in online advertising, our *first* objective is to predict the moment-to-moment zapping probability (MMZP) from users' facial expression during ad watching. The MMZP reflects users' dynamic interest level to the ad content. By collecting MMZP from diverse groups of users watching ads in multiple categories, our *second* objective is to extract user preference information from MMZP. To achieve these goals, we first describe a data collection paradigm and justify why "smile" is the expression of the central focus in analyzing users' feedback to an ad. Subsequently, we divide the smile responses into segments, which we term *responselet*, and learn a dictionary of responselet. Predictions of MMZP are made by classifying the smile responses using their sparse reconstruction coefficients as features. From the predictions of MMZP, rich knowledge about user preference is discovered, which is valuable for better understanding and promoting online ads.

2 Related Work

2.1 Automatic Facial Expression Recognition. Most of the facial expression recognition techniques can be divided into two groups, namely geometric-based approaches and appearance-based approaches. In the first group, the facial geometry over time is tracked and the type of expression is predicted based on the facial geometry deformation. Some of the popular methods include Active Shape Model (ASM) [9], Active Appearance Model (AAM) [12], particle filter [21], geometric deformation [10]. In the second group, the main idea is to extract robust feature representations to describe the appearance of facial features and their dynamics. For example, a bank of Gabor energy filters can be used to encode the facial texture [23], and Volume of Local Binary Patterns (VLBP) is also shown to be effective [28]. To achieve a global representation of a face without the loss of dynamics, Emotion Avatar Image (EAI) [26] has been proposed, which summarize the dynamics of the facial expression in a video into a single image. Some works combine both geometry and appearance for facial expression recognition [4]. For more details about advances in facial expression recognition, we direct the readers to a recent survey in [18].

2.2 Expression Analysis for Advertising Research. To study the spontaneous consumer facial expressions during ad watching, an interesting study was conducted by McDuff *et al.* [14] collected naturalistic facial response from people watching the US presidential debate in 2012. They found out that voter preference can be predicted with an average accuracy of over 73% based on facial expression responses. The AMFED dataset was introduced in [15], providing both spontaneous facial expression videos and self-report responses. Although this dataset does not concern users' zapping behavior, it provides a large testbed for "in-the-wild" consumer ad watching expression analysis. In a more realistic case, consumers may not be cooperative and can switch away from the viewing content at any moment. Thus, it is important and challenging for advertisers to retain users' attention during the playing of an ad. Elpers *et al.* [6] showed that the probability for an ad to be watched by users is affected by both the entertainment and the information value of the ad. Teixeira *et al.* [19] analyzed the user's zapping behavior using joy and surprise expressions. They found that the joy response highly correlates with user's zapping behavior. It was suggested that retaining a user's attention can produce desirable communication effects increase his/her interest as a potential consumer.

3 Experimental Design

3.1 Facial Expression Data. A facial expression dataset for online advertising analysis [25] is used in our experiments. In this dataset, 51 subjects were invited to watch multiple ads, each of which ranges from 30 to 90 seconds. Among those subjects, 16 are female and 35 are male. Regarding ethnicity, there are 9 Caucasians, 14 Asian Americans, 18 Asians, 8 Hispanics, and 2 African Americans. When watching an ad, a user can choose to finish watching or zap before it ends. In either case, 30 seconds were allowed for the user to neutralize their emotions. The selected 8 ads in our experiments cover two most popular categories: *Car* and *Fast Food*. The reason behind the choice of these two categories is that they have less gender bias as compared with ads in other categories such as *Beer* or *Makeup*. The selected ads in our experiments are all from mostly known brands, and these ads have different entertaining levels. Here, we use a scalar from 0 to 10 to rate the entertaining level of an ad from low to high. The scores for the ads were rated by an independent group of subjects and the averaged scores are used. Table 1 lists the ads information. In general the *Car* ads are more entertaining than the *Fast Food* ads with *Toyota* and *Nissan* being the two most amusing ones. All ads are available on primary websites such as YouTube. The playback order for these ads were randomized for each subject to diminish the ordering impact.

3.2 Motivation for Using Smile Response. Among common facial expressions, such as *fear* or *sadness*, smile response is leveraged in our method to predict zapping behavior due to the following motivations:

Table 1: Advertisements used in our experiments

Category	Brand	Ad Name	Entertaining Level	Length (in s)
Car	Toyota	I Wish	8	60
	Honda	We Know You	3	90
	Chevy	Wind Test	4	30
	Nissan	Enough	9	30
Fast Food	Jack In The Box	Hot Mess	5	30
	Subway	New Footlong	4	30
	Carl's Jr.	Oreo Ice Cream	6	32
	Pizza Hut	Make It Great	3	30

- It is shown by Elpers *et al.* [6] that the lack of *entertaining* factor in ads is a major reason for zapping. For the ad selection in this dataset, we focus on entertaining factor only. In addition, a large number of ads aim at amusing, so ads with various entertaining levels are widely available.

- In terms of facial expression response to ad watching, smile directly reflects the entertaining level of ads. It is observed that more than 95% of the detected expressions are smile in our dataset. A similar observation is also made recently [15].

- As an extensively studied facial expression, smile and its intensity can be accurately detected on a per frame basis and with simple features and a Support Vector Machine (SVM) classifier, smile classification can achieve an area under the curve (AUC) of 0.95 [27]. Examples of smile responses associated with zapping and non-zapping behavior are shown in Fig. 2.

3.3 Smile Intensity Estimation. To predict zapping behavior from smile response, we first explain how we estimate the smile intensity on a per frame basis. The face in each frame is first detected using Viola-Jones face detector [22]. The extracted faces are then aligned by the dense flow-based similarity registration technique [24]. Specifically, the registration is achieved by aligning every frame with a face to a reference face and the aligned faces are temporally smoothed. After alignment, the faces which are resized to 200×200 pixels. For feature extraction, each face is divided into non-overlapping blocks of 20×20 pixels in size. The Local Phase Quantization (LPQ) [16] features are extracted from each block, and features from all blocks are then concatenated to form the final feature representation for subsequent smile detection.

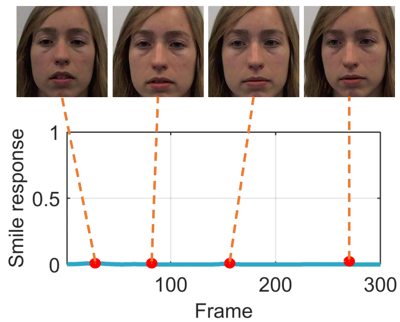
To detect smile, a binary classifier with the smiling face and neutral face being the two classes is trained. We employ the implementation of linear SVM in [3] to train the classifier. For accurate person-independent smile detection, we include a large number of subjects in training, from FEI dataset [20], Multi-PIE dataset [8], CAS-PEAL-R1 dataset [7], CK+ data [11], and images from Google search. As shown in [2], a combination of training data from different collections can improve the performance of the classifier. In total, 1543 smiling faces and 2035 neutral faces are included for training.

To validate the effectiveness of the smile classifier, we adopt a person-independent protocol, meaning that no test subject is included in the training data. Fig. 3(a) shows the Receiver Operating Characteristic (ROC) curve for 10-fold cross-validation. An Area Under Curve (AUC) of 0.98 is achieved, which indicates a high accuracy of the classifier. To further demonstrate the generalization ability of this classifier, we test a selection of 10,000 sample frames from the aforementioned online advertising dataset. The AUC in this case is 0.95 as shown in Fig. 3(b), which means that our smile classifier generalizes well on unseen data. The probability output of the SVM classifier is then recorded. As shown in [27], the smile intensity is correlated with the SVM probability output, and such probability output is used as a quantitative indicator of smile response in our experiments.

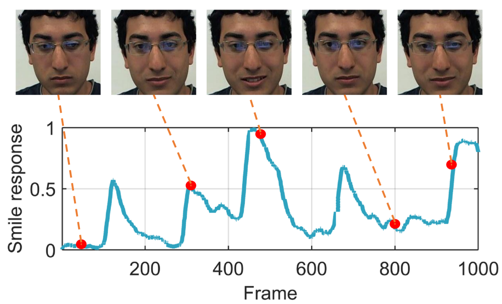
4 Methodology

To predict MMZP given smile responses, we formulate it as a zapping vs. non-zapping binary classification problem on a per frame basis. Below we explain in detail how zapping is quantified and how the features are extracted to predict zapping.

4.1 Zapping Quantification. Since the subjects are given the option to zap at any time, the fraction of an ad being watched varies for different participants. Fig. 4 shows the distribution of ad fraction that is being watched. It can be observed that most ads have been

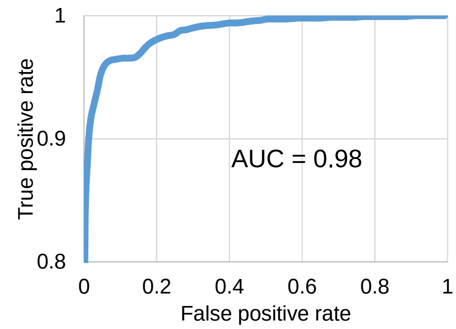


(a) Zapping

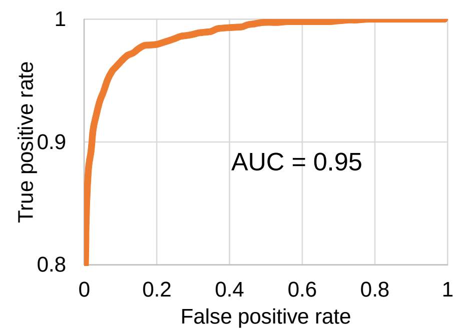


(b) Non-zapping

Figure 2: Examples of smile responses for zapping and non-zapping behavior. For the case of non-zapping, smile response intensity and dynamics are quite distinctive compared to the zapping case.



(a)



(b)

Figure 3: ROC curves of our person-independent smile detection from (a) 10-fold cross-validation on a collection of smiling and neutral faces and (b) 10,000 sample frames from the online advertising dataset [25].

watched at least 80% of their length. Besides, participants in our experiments tended to zap early if they are not attracted to an ad, as evidenced by the left part of the plot (i.e., 0% to 30% has a slightly higher probability). To distinguish zapped and non-zapped cases, a Gaussian mixture model with three components is fitted to the data. As seen in Fig. 4, the left-most and right-most Gaussian models are with low variance and well capture the distinctiveness of zapping and non-zapping class, respectively. The Gaussian model in the middle with high variance includes the sequences that are on the boundary of two classes. As a result, we empirically select the mean of the second mixture, 0.56, as the threshold to distinguish zapping from non-zapping sequences.

4.2 Zapping Prediction. During the training phase, a dictionary of *responselet* (i.e., small segments of smile response) is learned from the recorded subjects' facial responses. Then given a smile responselet, it is first reconstructed by the learned dictionary and the reconstruction coefficients are used as features to train a zapping prediction model. During testing, reconstruction coefficients for a responselet is used for zapping prediction with the trained model.

Responselet Dictionary Learning. The history of the dictionary design can be traced back to the fast Fourier transform (FFT) or wavelets. In recently years, dictionary learning for sparse representation has prevailed in signal processing and computer vision tasks [13]. In our case, we learn the dictionary which can well reconstruct the input signal in a sparse manner.

Concretely, a smile responselet is defined as a 1-D time series data sample of length m, whose value ranges from 0 to 1. Given a set of n smile responselet, $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathcal{R}^{m \times n}$, a dictionary $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_p] \in \mathcal{R}^{m \times p}$ is learned by minimizing

$$\min_{\mathbf{D}, \boldsymbol{\alpha}} \sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 + \lambda ||\boldsymbol{\alpha}_i||_1, \tag{4.1}$$

where $\forall j = 1, \cdots, k, \mathbf{d}_j^\top \mathbf{d}_j \leq 1$. λ is a regularization parameter to control the sparsity on $\boldsymbol{\alpha}$. The ℓ_2 regularization term prevents arbitrarily large values of $\mathbf{D}$, and

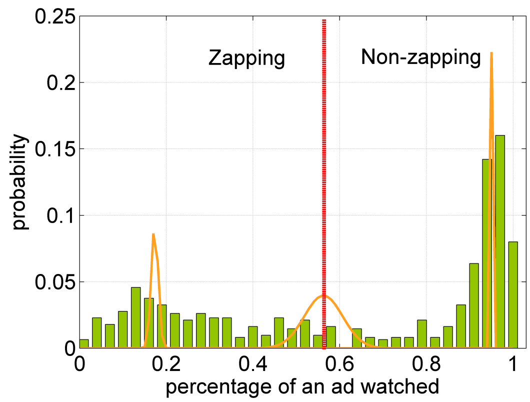


Figure 4: The zapping distribution. We use Gaussian mixture model to fit the data. The left-most and right-most Gaussians well capture the distinctiveness of zapping and non-zapping classes, respectively. The middle Gaussian with high variance represents the sequences that are on the boundary of two classes. We use the mean of the second Gaussian, 0.56, as a data-driven threshold to distinguish zapping from non-zapping.

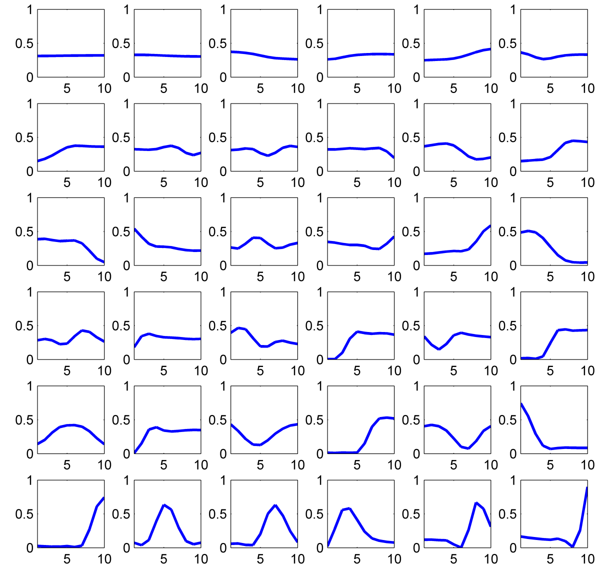


Figure 5: Visualization of the learned responselet dictionary

the ℓ_1 regularization term results in a sparse solution for $\boldsymbol{\alpha}$. Eq. (4.1) can be solved efficiently by an online dictionary optimization algorithm [13].

Consider the facial response of a subject watching an ad as one sequence, with 51 subjects and 8 ads, there are 408 sequences. As seen in Fig. 2, the dynamics of the smile response has different lengths. Therefore, each sequence is normalized to 100 frames by interpolation or extrapolation. Each responselet is extracted from $m = 10$ consecutive frames, and the stride of the moving window for the next responselet is set to one frame. As a result, approximately $40,000$ responselets are extracted from all sequences. A dictionary with $p = 36$ responselet atoms is learned with each atom visualized in Fig. 5, sorted by their absolute gradient variance in an ascending order.

As seen in Fig. 5, the learned dictionary atoms start from flat shape (upper left), gradually vary in slope, and finally transform to shapes representing much more drastic changes (bottom two rows) indicated by the bell shape. The wide variety of the learned dictionary ensures that a responselet can be well represented by the dictionary.

Training of Zapping Prediction Model. Having learned the dictionary $\mathbf{D}$, with k frames in a sequence, we first divide the response from frame 1 to k into $k - 10 + 1$ number of responselets (i.e., 10 frames per responselets) with a stride of one frame in the same way as aforementioned. Then, the sparse coefficients $\boldsymbol{\alpha}_i$ for

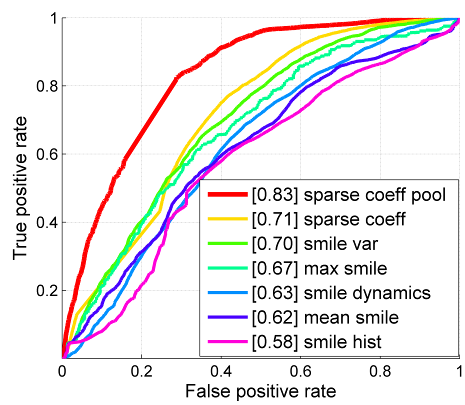


Figure 6: The ROC curves for zapping classification using various features. Numbers are the AUC values.

each responselet is obtained by

$$\min_{\boldsymbol{\alpha}_i} ||\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 + \lambda||\boldsymbol{\alpha}_i||_1. \tag{4.2}$$

Thus, for each responselet, we obtain a $p \times 1$ reconstruction coefficients vector. To encode the between frame information, we perform mean-pooling of all available coefficient vectors for the responselets generated from frame 1 to k, yielding our final feature representation of size $p \times 1$.

Zapping Classification. In order to distinguish zapping from non-zapping sequences, we formulate it as

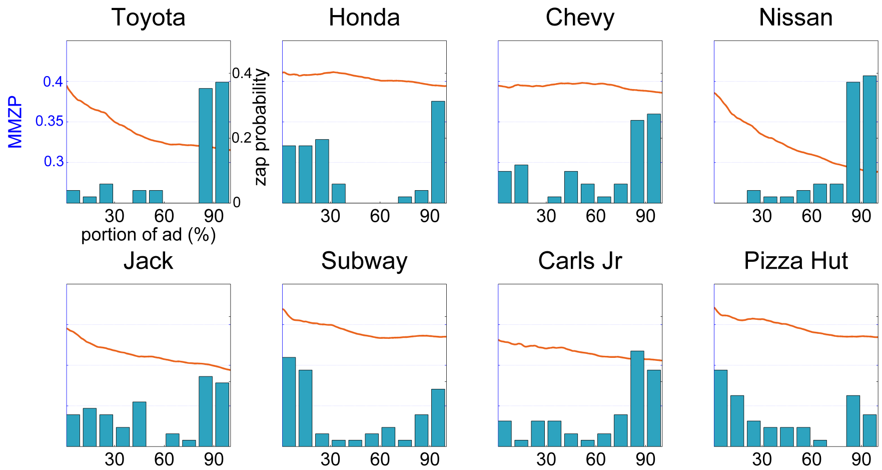


Figure 7: MMZP of different ads. The lines show the MMZP trends and the bars show the corresponding ground truth zapping distribution. The axis and labels are the same across all sub-figures and are displayed in the first panel only for compactness. We use the same display settings for the rest of the figures.

a binary classification problem. The class labels of the data are assigned based on empirically determined threshold shown in Fig. 4.

During training, at each moment, the features extracted from responselet are used to train a linear SVM classifier with its class label determined by the label of the entire sequence. The zapping classifier is evaluated by leave-one-subject-out cross-validation. In each fold, one subject is held out for testing, and the others are used for training. Besides mean-pooled features, we have also included several other features as baselines for comparison, including: *mean*, *max*, *variance of smile response*, *smile dynamics* (meaning that we directly use the length-normalized smile response as features), and *smile histogram* [27]. In addition, direct use of the sparse coefficients without mean-pooling is compared as a baseline.

As shown in Fig. 6, sparse coefficients with mean-pooling significantly outperforms the other features. This is attributed to the fact that mean-pooling takes into the account the between-frame information and yields a more stable representation than other features. Smile variance performs slightly better than other smile response variants. This also demonstrates that the shape of a smile response is discriminative to distinguish zapping from non-zapping.

During testing, features at each moment are extracted in the same manner as training, and then passed to the zapping classifier. The probability output is considered as our final MMZP representation. In this manner, we are able to predict the zapping probability at each moment. In the following section, whether the MMZP is a valid metric is analyzed first. Then, several interesting observations discovered from the data regarding the effect of gender, ethnicity, and personal preference are presented by analyzing the MMZP.

5 Knowledge Discovery

As aforementioned, MMZP represents the dynamics of users' interest level while watching an ad. In general, increased MMZP correlates with more chances of zapping due to lack of interest in the ad content. To validate this conclusion, we first examine the effectiveness of MMZP, and then provide a series of knowledge discovery from the observations on MMZP changes.

MMZP Validation. To validate the accuracy of MMZP, we first compare it with the actual user zapping behavior. In Fig. 7, the mean MMZP (left axis) of all subjects for each ad is plotted and is compared with the ground truth zapping behavior as shown in the bar plot of each panel (right axis). Note that the zapping behavior is represented by the percentage of an ad being watched, which was recorded during data collection. In general, the fast decreasing pattern of MMZP (less likely to zap) correlates with zapping behavior occurring near the end of an ad (e.g., *Toyota* and *Nissan*). In other words, zapping near the end suggests that more interesting contents in the ad were found by the user. On the other hand, slowly decreasing or stable MMZP indicates more chances of zapping (e.g., *Chevy* and *Pizza Hut*). This is congruent with both zapping distribution in the bar plot and the ad entertaining level of each ad as listed in Table 1.

Gender Preference for Ads. We consider both gender groups watching different ads and discover interesting findings illustrated by some examples in Fig. 8. Based on the MMZP changes, it can be observed that: 1) *Females* prefer *Honda*, probably because of the embedded story about family and pet along with gentle background music; 2) *Males* favor *Jack in the Box*, most likely due to the rock music scene in the ad. Similar observations hold for other ads and these findings suggest that during ad design gender preference shall be leveraged for more precise targeting.

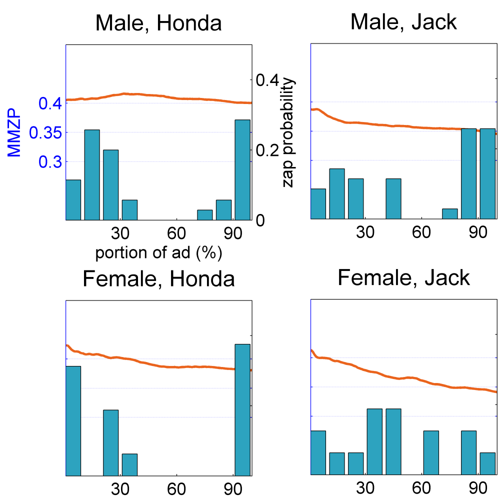


Figure 8: Gender-related ad preference. The *Male* group prefers *Jack in the Box* in which rock music scene is shown. The *Female* group enjoys *Honda* more likely due to its family and pet story setting with gentle background music.

Ethnicity Preference for Ads. As seen in Fig. 9, *Asians* show less interest in an entertaining *Toyota* ad. We found out that this is because *Toyota* ad's entertaining factor comes from amusing conversations; while some members in the Asian group, who have just come to the US with English not as their first language, did not catch the full meaning of the conversation. Thus, more zapping happens for the *Asian* group. For another example, *Hispanic* group is less interested in a more direct, food scene oriented *Pizza Hut* ad. Our interview with them suggests that most of them retain their ethnic dining habit in which pizza plays a less important role. Such examples indicate that ad preference is related to a user's ethnicity group and taking this into account in ad design and targeting may result in a better ad campaign.

Personalized User Preference for Ads. Targeted

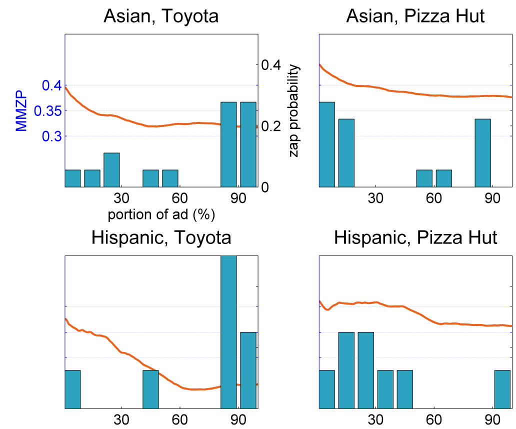


Figure 9: Ethnicity-related ad preference. It is observed that *Asian* group is more likely to zap on *Toyota* while *Hispanic* group does not much enjoy *Pizza Hut*.

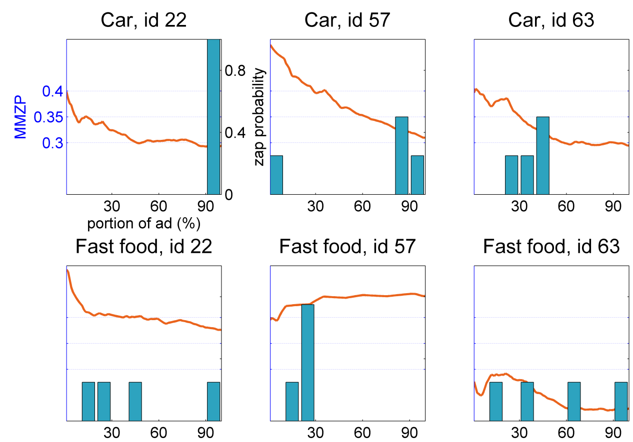


Figure 10: Examples of personal zapping behavior. It is observed that subject 22 and 57 prefer *Car* ads while subject 63 likes *Fast Food* ads better.

advertising has proven to be effective in online marketing, as evidenced by the dissemination of online personalized recommendations. We argue that the proposed MMZP can also be used as a user preference metric and therefore, has the potential to help with better personalized advertising. In Fig. 10 we show three subjects' MMZP from watching ads in both categories. The decreasing MMZP towards the end of an ad for subject 22 and 57 suggests that they prefer *Car* ads more than *Fast food* ads in which their MMZP remains high across the sequence. On the other hand, based on the changes in MMZP, subject 63 is more interested in *Fast food* than *Car*. These observations are also confirmed by the corresponding ground truth zapping distribution.

Thus, using MMZP can provide a better understanding of personal zapping behavior, which is essential for more accurate personalized ad targeting.

6 Conclusions

Online user behavior analysis plays an important role in marketing and advertising. In this paper, we have proposed a metric termed moment-to-moment zapping probability (MMZP) for user behavior modeling and understanding based on users' facial expressions. Studies showed that MMZP can be predicted using the smile responses, and knowledge regarding user behavior and preference can be discovered from MMZP. The findings by analyzing the MMZP may facilitate better group-specific or even person-specific advertisement design and targeting.

References

[1] A. Abbasi, R. Y. Lau, and D. E. Brown. Predicting behavior. *IEEE Intell. Syst.*, 2015.

[2] L. An, S. Yang, and B. Bhanu. Efficient smile detection by extreme learning machine. *Neurocomputing*, 2014.

[3] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2011.

[4] J. Chen, D. Chen, Y. Gong, M. Yu, K. Zhang, and L. Wang. Facial expression recognition using geometric and appearance features. In *ICIMCS*, 2012.

[5] P. Ekman. Facial expression and emotion. *American Psychologist*, 1993.

[6] J. L. W. Elpers, M. Wedel, and R. G. Pieters. Why do consumers stop viewing television commercials? Two experiments on the influence of moment-to-moment entertainment and information value. *J. Marketing. Res.*, 2003.

[7] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The CAS-PEAL large-scale chinese face database and baseline evaluations. *IEEE Trans. Syst., Man, Cybern. A, Syst.,Humans*, 2008.

[8] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image Vision Comput.*, 2010.

[9] C. Hu, Y. Chang, R. Feris, and M. Turk. Manifold based analysis of facial expression. In *CVPR*, 2004.

[10] I. Kotsia and I. Pitas. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Trans. Image Process.*, 2007.

[11] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *CVPR*, 2010.

[12] S. Lucey, I. Matthews, C. Hu, Z. Ambadar, F. De La Torre, and J. Cohn. AAM derived face representations for robust facial action recognition. In *FG*, 2006.

[13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.*, 2010.

[14] D. McDuff, R. E. Kaliouby, E. Kodra, and R. W. Picard. Measuring voter's candidate preference based on affective responses to election debates. In *ACII*, 2013.

[15] D. McDuff, R. E. Kaliouby, T. Senechal, M. Amr, J. F. Cohn, and R. W. Picard. Affectiva-MIT facial expression dataset (AM-FED): Naturalistic and spontaneous facial expressions collected in-the-wild. In *CVPR Workshops*, 2013.

[16] V. Ojansivu and J. Heikkilä. Blur insensitive texture classification using local phase quantization. In *ICISP*, 2008.

[17] K. Poels and S. Dewitte. How to capture the heart? Reviewing 20 years of emotion measurement in advertising. Technical report, Katholieke Universiteit Leuven, 2006.

[18] E. Sariyanidi, H. Gunes, and A. Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015.

[19] T. Teixeira, M. Wedel, and R. Pieters. Emotion-induced engagement in internet video advertisements. *J. Marketing. Res.*, 2012.

[20] C. E. Thomaz and G. A. Giraldi. A new ranking method for principal components analysis and its application to face image analysis. *Image Vision Comput.*, 2010.

[21] M. Valstar, I. Patras, and M. Pantic. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. In *CVPR Workshops*, 2005.

[22] P. Viola and M. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 2004.

[23] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan. Toward practical smile detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009.

[24] S. Yang, L. An, B. Bhanu, and N. Thakoor. Improving action units recognition using dense flow-based face registration in video. In *FG*, 2013.

[25] S. Yang, L. An, M. Kafai, and B. Bhanu. To skip or not to skip? A dataset of spontaneous affective response of online advertising (SARA) for audience behavior analysis. In *FG*, 2015.

[26] S. Yang and B. Bhanu. Understanding discrete facial expressions in video using an emotion avatar image. *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, 2012.

[27] S. Yang, M. Kafai, L. An, and B. Bhanu. Zapping index:using smile to measure advertisement zapping likelihood. *IEEE Trans. Affect. Comput.*, 2014.

[28] G. Zhao and M. Pietikäinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007.