Zapping Index:Using Smile to Measure Advertisement Zapping Likelihood

Songfan Yang, *Member, IEEE*, Mehran Kafai, *Member, IEEE*,
Le An, *Student Member, IEEE*, and Bir Bhanu, *Fellow, IEEE*

Abstract—In marketing and advertising research, "zapping" is defined as the action when a viewer stops watching a commercial. Researchers analyze users' behavior in order to prevent zapping which helps advertisers to design effective commercials. Since emotions can be used to engage consumers, in this paper, we leverage automated facial expression analysis to understand consumers' zapping behavior. Firstly, we provide an accurate moment-to-moment smile detection algorithm. Secondly, we formulate a binary classification problem (zapping/non-zapping) based on real-world scenarios, and adopt smile response as the feature to predict zapping. Thirdly, to cope with the lack of a metric in advertising evaluation, we propose a new metric called Zapping Index (ZI). ZI is a moment-to-moment measurement of a user's zapping probability. It gauges not only the reaction of a user, but also the preference of a user to commercials. Finally, extensive experiments are performed to provide insights and we make recommendations that will be useful to both advertisers and advertisement publishers.

Index Terms—Online advertising, smile detection, Zapping Index (ZI), user preference

1 Introduction

In recent years, multimedia data (e.g., images, videos, audios) on the Internet keep on increasing at a phenomenal rate. For example, 72 hours of video data are uploaded to the YouTube every minute [1]. More and more people tend to spend time watching videos on the Internet instead of using the traditional media such as TV. In addition, with the vastly growing popularity of mobile devices (e.g., smart phones, tablets), easy Internet access continues to attract more traffic on mobile networks. As predicted, in 2017 the video contents will account for 66 percent of all mobile data traffic.[1]

The popularity of the Internet videos implies a huge potential for online commercial advertisement (ad). The marketing expenses for commercial ads on the Internet are growing. For instance, the cost of a 30-second commercial on TV at prime time in the US was around 0.5 million US dollars in the fall of 2012.[2] At some specific venues, the cost of commercials may be much higher. As an example, the cost of a 30-second commertial in the Super Bowl event in US has hit 4 million US dollars in 2013.[3] With the increased advertising cost on TV and decreased audiences, marketers are gradually switching their focus to online advertising, in favor of their large audience base and lower cost to publish.

As a well-known example, the TrueView in-stream advertising [2] is a popular online advertising tool by Google Inc. The ad is shown prior to the video requested by the user. The user has the option to skip the ad and move directly to the desired video after 5 seconds of viewing the ad. The advertisers are billed if a user watches the ad at least for 30 seconds or the complete ad, if it is less than 30 seconds long. In such a case, for the online media provider (e.g., Google) to obtain the maximum profit and for the advertiser to reach the widest audience and achieve the advertising goal, it is their common interest to draw viewers' attention to the online commercials.

In marketing and advertising research, to evaluate the attention to the commercials, zapping is considered as an important topic [3], if not the most important one. Commercial viewers often have the option to 'zap' a commercial by either switching channels or simply turning off the source. The action of zapping indicates that the viewer is no longer interested in the commercial and this behavior means the loss of a consumer for the advertiser. To evaluate the effectiveness of advertising, several methods can be adopted. Self-report, which registers a respondent's subjective feeling, suffers from an important limitation referred as "cognitive bias", and may not always be able to capture lower-order emotions in an accurate way [4].

Facial expression is one of the richest source of communication [5]. Automatic facial expression recognition finds its applications in human behavior analysis, human-human interaction and human-computer interaction. Automatic facial expressions analysis is non-intrusive and can be dynamically analyzed as a commercial is playing [6]. Accurate facial expression analysis facilitates the marketing and advertising researchers in understanding a user's emotional

1. http://www.cisco.com/en/US/netsol/ns827/networking_solutions_sub_solution.html.
2. http://domainestimations.com/?p=14174.
3. http://www.forbes.com/sites/alexkonrad/2013/02/02/even-with-record-prices-10-million-spot/.

- *S. Yang, L. An, and B. Bhanu are with the Center for Research in Intelligent Systems, University of California, Riverside, CA 92521. E-mail: {syang, lan, bhanu}@cris.ucr.edu.*
- *M. Kafai is with Hewlett Packard Laboratories, Palo Alto, CA 94304. E-mail: mehran.kafai@hp.com.*

Manuscript received 5 Nov. 2013; revised 26 Aug. 2014; accepted 13 Oct. 2014. Date of publication 23 Oct. 2014; date of current version 2 Dec. 2014. Recommended for acceptance by Q. Ji.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TAFFC.2014.2364581

state and behavior. This has the potential to improve the effectiveness of advertising or even design interactive commercials to enhance the advertising experience.

Recently, smile has been demonstrated as an useful indicator of a user's preference of commercials [7]. Teixeira et al. [6] develop a statistical approach using facial expressions to study advertising and they find that surprise and joy are effective in retaining a viewer's attention. Furthermore, applying machine learning and data mining techniques to advertising research enables us to exploit the underlining relationships between commercials and users by performing experiments with large data.

In this paper, we attempt to understand a user's behavior in watching an ad. We make prediction on a user's zapping probability and provide guidance to ad publishers and advertisers. This can benefit ad publishers (such as YouTube) to understand the user's reaction to a certain commercial and, therefore, decide its value. Besides, this can also benefit advertisers so that they have an evaluation tool to analyze the feedback of their ad. Advertisers can leverage this behavior feedback to make better commercials.

We propose a measurement called Zapping Index (ZI), which is a prediction of the moment-to-moment zapping probability when an user is watching a commercial. The motivation for developing ZI is the following:

- The need for marketing metrics is well recognized. A survey of CEOs shows that CEO's top concern about marketing was the lack of performance metrics [8]. ZI creates a new metric for marketer and advertisers.
- ZI helps to study the affective behavior of an audience.
- ZI helps to improve the effectiveness of an ad.

To calculate the ZI, we opt to use smile response and set it apart from other facial expressions. As demonstrated in [3], entertaining information has a strong relation to zapping. Smile is a reflection of joy and happiness triggered by entertainment. Moreover, current computer vision algorithms perform well on automatic smile detection [9], [10].

The rest of the paper is organized as follows: Section 2 reviews the existing methods for automated human facial expression recognition and zapping analysis. In Section 3, after introducing our accurate smile detection, we illustrate the data collection procedure and the data characteristics, which will motivate the selection of features for zapping detection/classification. Section 4 provides a series of experiments, demonstrating the effectiveness of the proposed Zapping Index. Finally, Section 5 concludes the paper.

2 Related Work and Our Contributions

2.1 Automatic Facial Expression Recognition

Facial expression recognition techniques can be broadly divided as geometric-based approaches, appearance-based approaches, and the combination of both.

Geometric-based approaches track the facial geometry over time and infer expression based on the facial geometry deformation. Some exemplar methods include: Active Shape Model (ASM) [11], Active Appearance Model (AAM) [12], particle filter [13], geometric deformation [14]. Appearance-based approaches, on the other hand, emphasize on describing the appearance of facial features and their dynamics. Whitehill et al. [10] use a bank of Gabor energy filters to decompose the facial texture. The volume of local binary patterns (VLBP) is extracted in [15]. Yang and Bhanu [16] aggregate the dynamics of the facial expression into a single image, Emotion Avatar Image (EAI), for high accuracy person-independent expression recognition.

2.2 Zapping Analysis

The attention paid to a commercials determines the interests of the audience in that commercial and only those commercials that retain a viewer's attention can produce desired communication effects [6]. As the consumers have a choice to switch away from either a TV commercial or an online video commercial, it is challenging for the advertisers to retain consumers' attention during the course of a commercial [3]. The term zapping implies that the receiver of a commercial is no longer interested in its content/presentation, thus opt not to continue watching the commercial. In [17] a hierarchical Bayes approach is used to analyze the dynamics of attention to TV commercials. It investigates how the likelihood of a commercial zapping varies with time and shows that across-ad heterogeneity in zapping is related to the underlying characteristics of the commercial. Elpers et al. [3] demonstrate that both the entertainment and the information value of a commercial have a strong multiplicative effect on the probability for a commercial to be watched by viewers. Two experiments with a total number of 190 subjects and 45 commercials were conducted to support this finding. Kooij et al. [18] show that zapping has influence on end-user's Quality of Experience (QoE) for Internet Protocol television (IPTV). Further study on zapping is conducted in [19] and various solutions are proposed to reduce zapping to keep the user staying with the IPTV broadcasting. Teixeira et al. [6] incorporate joy and surprise expression recognition from a Bayesian Neural Network classification system to analyze the user's zapping decision. They conclude that the velocity of the joy response highly impacts the viewer's zapping behavior. But they have not made any prediction on the moment-to-moment zapping probability.

2.3 Our Contributions

The contributions of this work are summarized as follows:

1) We introduce an accurate person-independent video-based smile detection method. The smile response intensity is well transformed from a probability score from 0 to 1.
2) We perform zapping detection/classification in a non-intrusive manner based on facial expression cues.
3) We propose a novel metric called Zapping Index for ad evaluation. ZI is a moment-to-moment prediction of a user's zapping probability.
4) We collect a database, named AdEmotion, for the analysis of zapping prediction and user preference evaluation. We demonstrate the usefulness of ZI in measuring user preference. This results in advices for both ad publishers as well as providers about the effectiveness of an ad. The AdEmotion dataset will be publicly available in the near future on our website.

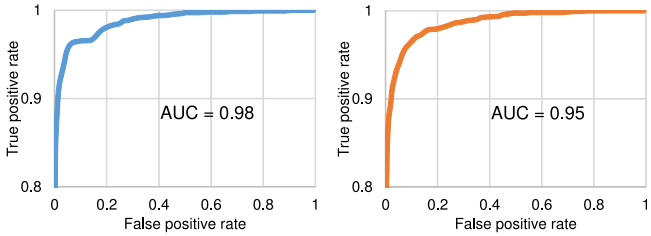


Fig. 1. ROC curve for our person-independent smile detection algorithm.

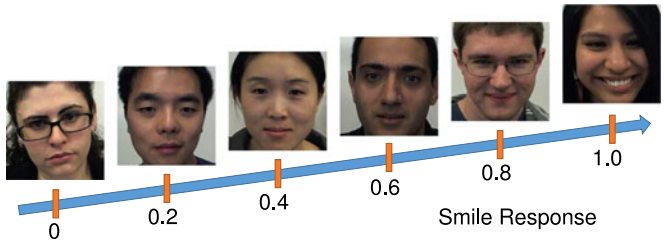


Fig. 2. Sample smile response results. The response value reflects the intensity of smile.

3 Technical Approach

In this section, we first introduce how moment-to-moment smile detection is carried out. Subsequently, we describe the data collection procedure. We formulate the problem of distinguishing zapping from non-zapping as a binary classification problem. After analyzing the characteristics of the data, we propose a new feature which is a temporal histogram of the smile measurement over time. We adopt SVM classifier to train the zapping classification model, which is then used to generate the Zapping Index.

3.1 Smile Detection

The goal is to compute the probability of smile on a per-frame basis. The faces are first extracted using Viola-Jones face detector [20]. We then follow our previous work [21] to align the faces using dense flow-based similarity registration technique. This registration algorithm aligns every frame with a face to a reference face and the registration results are temporally smoothed. Thus, the person-independent spontaneous facial expression recognition can be carried out in a meaningful manner. The aligned faces which are scaled to 200×200 pixels, are divided into 20×20 pixel regions. The Local Phase Quantization [22] texture descriptor is computed for each of the regions. These outputs are then concatenated to form the feature for smile detection.

The smile detection is formulated as a binary classification problem with the smiling face and neutral face being the two class labels. We adopt the linear Support Vector Machine (SVM) [23] for classification. For accurate person-independent smile detection, the classifier is trained on multiple databases with a large number of subjects from: FEI [24], Multi-PIE [25], CAS-PEAL [26], CK+ [27], and data from Google image search similar to [28]. In total, 1,543 subjects (1,543 smiling faces and 2,035 neutral faces) are included for training.

A series of tests are carried out in a person-independent manner where no test subject is included during training. The Area Under Curve (AUC) is 0.98 for the 10-fold cross validation (see Fig. 1a). To demonstrate the generalization of this classifier, we carried out a test on a selection of 10,000 sample frames from our AdEmotion database (see Section 3.2) that we collected in this research. The Area Under Curve is 0.95 in Fig. 1b, which means that the smile classifier performs well on unseen data. The probability output of the SVM smile classifier is then recorded as the smile response. Since we structure *smiling* vs. *neutral* instead of *smiling* vs. *non-smiling* classification, an interesting finding is that, the classification rate of test data is not only superior, but also the probabilistic output of smile detector is able to capture the smile intensity as illustrated in Fig. 2. The reason we choose smile vs. neutral expression classification setup in this experiment is that the subjects are concentrated on the viewing experience. Most of the expressions other than smile are of neutral nature, and very few subjects display excessive non-smile expression. In this case, the probabilistic outputs closely correlate to the smile intensity even when it is low. One thing worth noting is that, there are neutral examples with open mouth in the training data, and therefore, the classifier is not just naively predicting random mouth motion but rather muscle motion caused by smile.

For proof of concept, we have verified the probabilistic outputs with the manually annotated smile intensity results. We have gathered three annotators, and each is given 500 frames sampled from the entire AdEmotion data. The annotators score the smile intensity of each frame by comparing it with the reference figure similar to Fig. 2. The median value of all three annotators is selected as the ground-truth smile intensity to mitigate the effect from a large discrepancy among annotators. The resulting absolute mean error intensity is 0.216 between the prediction and ground-truth.

Some failure cases are shown in Fig. 3a. In order to eliminate the subjects whose smile response is inaccurate, we leverage the fact that the expression for a subject is distributed around the neutral as we mentioned earlier. Therefore, for each sequence, we quantize the smile response to 0.1 accuracy, and take the *mode* of the quantization to approximate the baseline expression response for a subject. As a result, all the seven error cases, whose smile baseline is 1, are able to be separated as shown in Fig. 3b.

3.2 AdEmotion Data Collection

Participants were seated in front of a 23 inch monitor, with a Logitech c910 webcam mounted on the top of the

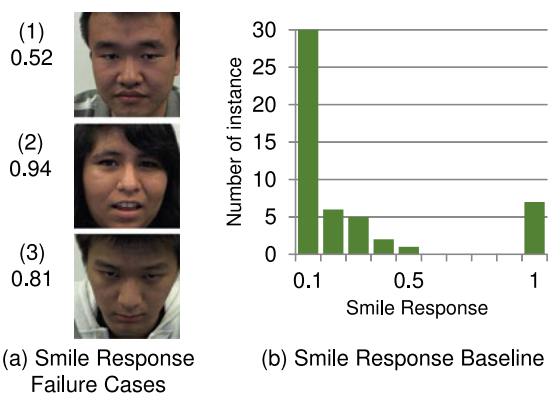


Fig. 3. Handling smile detection failure. (a) Smile response failure cases where (1) and (2) exist because the subjects' neutral faces are similar to smiling faces; (3) is due to out-of-plane rotation. (b) The baseline response of each subject is used to eliminate the failure cases automatically.

TABLE 1
Advertisement Selection

Category	Brand	Ad Name
Car	Toyota	I Wish
	Honda	We Know You
	Chevy	Wind Test
	Nissan	Enough
Fast Food	Jack In The Box	Hot Mess
	Subway	New Footlong
	Carl's Jr.	Oreo Ice Cream
	Pizza Hut	Make It Great

monitor. The webcam resolution is set to 960×544 pixels. The average resolution on face is approximately 220×220 pixels. Participants were shown a series of eight video ads in random order selected from two categories shown in Table 1. The length of the ad ranged from 30 to 90 seconds. Participants were instructed that they could watch each ad until the end or zap at any moment by clicking on the skip button. In either situation, participants were given a 30 seconds break to reduce the emotional effect to the subsequent ad watching experience. They were also given a questionnaire during each break that contained the following questions:

1) Did you like the commercial?
2) Did you skip the ad?
3) Why did you skip? Mark all that applies:
   - The ad is not funny.
   - The ad is not informative.
   - I have seen this ad before.

There has been research [3] that shows that the lack of entertainment and information factors are the two major reasons for zapping. We design these questions in order to analyze different aspects of this dataset. In this work, our focus is on using the response provided by subjects to verify the predictions from the zapping index, which will be discussed later in detail.

The entire data collection procedure lasts 8 minutes on the average for each participant, and no one is interrupted during this procedure. The participants' facial behavior during the entire procedure is recorded by the webcam at 24 fps.

During recording, there is a secondary monitor behind the participant, which displays the same content watched by the participant. The recording camera is able to capture a subject's facial expression as well as the corresponding content that he or she is watching. We designed this setting for data synchronization. In order to analyze the facial expression responses of different participants with respect to certain ad, we manually separated the expression data according to the ad information shown on the duplicate monitor. No data during the interval of two ads is used in this work. The setup of data collection environment is shown in Fig. 4.

The ads shown in Table 1 are selected by the following criteria:

1) Popular category. The *Fast Food* and *Car* are the two categories that almost everyone is familiar with and well connected to in the United States.

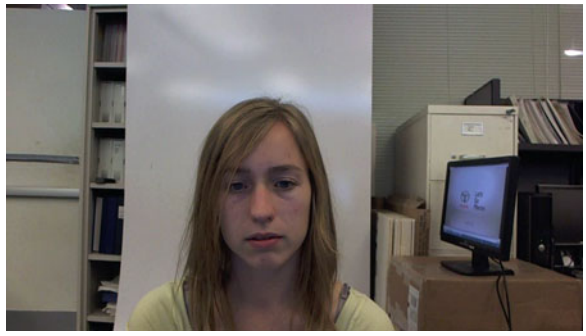


Fig. 4. The data collection environment. A duplicate monitor is used for data synchronization.

2) Minimum gender bias. We do not consider the gender effect in this research. The selected ad categories have less gender bias compared with categories such as *Beer* or *Makeup*.
3) Recognizable brand. Since online video user is the target market. We select the ad from brands that either have their official YouTube Channel or participated in the YouTube ad campaign. In this way, we have access to the ad for this research.
4) Varying entertainment levels. We have carefully evaluated the entertainment information in each ad. Our final ad selection includes both kinds of ads that are very amusing and that are less entertaining.

Fifty five college students have participated in our data collection. There are 31 percent female, 40 percent Asian, 25 percent Euro-American, 16 percent Afro-American, and 19 percent other ethnicity groups.

3.3 Data Characteristics

We analyze the characteristics of AdEmotion dataset in terms of zapping distribution and smile response. These characteristics are essential in motivating our zapping classification feature. We could potentially design systems that recognize other facial expressions. However, in this application, subjects concentrated on watching ads, and therefore, the dominant facial expressions are neutral and smile. This is also demonstrated to be true in an "in-the-wild" ad-watching data, namely AMFED [29]. Therefore, in light of the idea of Occam's razor, we have focused specifically on the smile expression specifically in this work.

3.3.1 *The Zapping Distribution*

Since participants are given the option of zapping at anytime, we show the distribution of fraction of an ad that is being watched in Fig. 5. In other words, it is the distribution of the portion of the ad that has been watched. We fit a Gaussian mixture model with two components to the distribution and find that 90 percent of the ad fraction is the best value to separate the two components of the mixture. In Fig. 5, the probability is dramatically higher in 90 to 100 percent range. This means that a large portion of the ads have been watched until the end. In the 0 to 90 percent range, the first half (0 to 45 percent) has a slightly higher probability than the second half on the average. This informs us that subjects in our experiments tend to zap early if they do not feel like watching an ad.

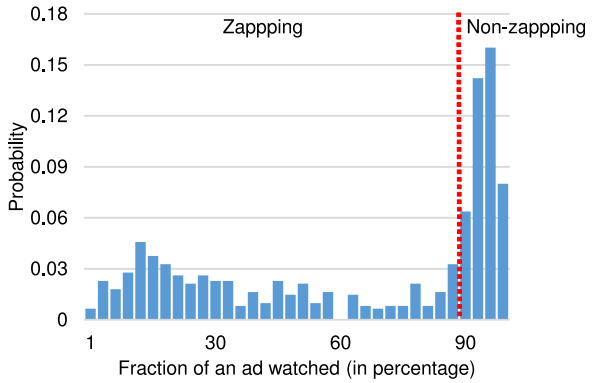


Fig. 5. The zapping distribution. The data-driven threshold at 90 percent is used to separate the data into zapping and non-zapping classes.

One interesting fact that is worth noting is that the popular TrueView advertisement publisher only bills the advertiser if an ad has been watched for more than 30 seconds [2]. If we have a better understanding of the zapping behavior, we can create a win-win-win situation: the user receives more desirable video content; the advertiser obtains more attention from users; and the publisher (such as YouTube from Google) gains more revenue.

Thus, based on the zapping distribution (see Fig. 5), we define two different classes: zapping and non-zapping, and use 90 percent of the ad length as the threshold in separating these classes. Given the facial expression response of a user watching an ad, one of our goals is to determine in an automated manner the class of the sequence. This can be formulated as a binary classification problem where zapping is the positive class and non-zapping is the negative class. The analysis of the class characteristics in Sections 3.3.2, 3.3.3, and 3.3.4 provides us the motivation for the feature used in zapping classification.

3.3.2 The Mean Smile Response

We conduct person-independent smile detection as described in Section 3.1. We present our motivations to the feature selection for zapping classification, and ultimately, establish a strong correlation of the zapping index from our predictor and the viewer's zapping behavior.

We analyze the average smile response in the first 30 seconds (720 frames from our 24 fps webcam device) for both zapping and non-zapping classes. In Fig. 6, the moment-to-moment mean smile response is bounded by the positive and negative standard error of the mean (SEM). Since the user can zap at any time, the sequence are of various lengths. Therefore, the average smile response is computed as follows:

$$r_m(t) = \frac{\sum_{i=1}^{N} r_i(t)}{\sum_{i=1}^{N} I_i(t)}, \; I_i(t) = \begin{cases} 1, & \text{if } r_i(t) \text{ exists,} \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where $r_i(t)$ is the smile response of sequence i at time t, N is the number of sequences, $I_i(t)$ is the indicator function to check the existence of the smile response of sequence i at time t. In other words, for each frame, the average smile response is computed based on the available responses. Using the similar idea, we compute the SEM bound by

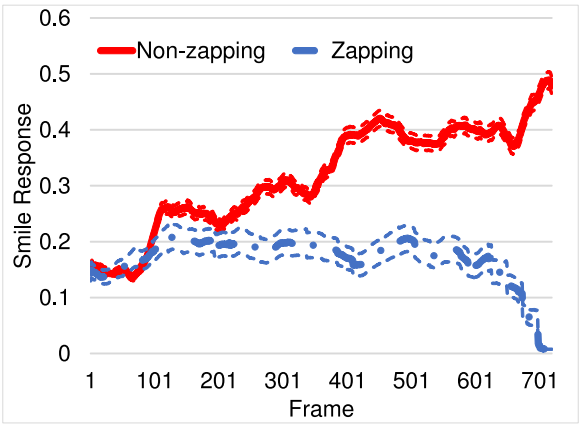


Fig. 6. The average smile response of zapping and non-zapping classes for the first 30 seconds (720 frames). Each sequence is bounded by its standard error of the mean.

$$sem(t) = \frac{std(r(t))}{\sum_{i=1}^{N} I_i(t)} = \sqrt{\frac{\sum_{i=1}^{N} (r_i(t) - r_m(t))^2}{\sum_{i=1}^{N} I_i(t)(\sum_{i=1}^{N} I_i(t) - 1)}}, \quad (2)$$

where $r_m(t)$ and $std(r(t))$ are the mean and the standard deviation of available smile responses at time t, respectively.

In Fig. 6, the smile response level for the two classes is initially about the same. Thereafter, the response of the non-zapping class increases for the rest of the 30 seconds. On the contrary, for the zapping class, the response remains around 0.2 and decreases toward the end. *Therefore, the moment-to-moment average smile response is a good feature to separate zapping from non-zapping class.* This observation is also in line with the conclusion in [6] that smile level largely correlates with the zapping behavior.

3.3.3 The Maximum Smile Response

The maximum smile response of the sequences is also different for zapping and non-zapping classes. Two examples are shown in Fig. 7.

We plot the distribution of sequences from the two classes based on their maximum smile response in Fig. 8. The total probability of each group sums up to 1. As illustrated in Fig. 8, if a sequence's maximum smile response is above 0.5, then the chance is higher that it belongs to the non-zapping class, and vice versa for maximum smile response below 0.5. The probability reaches the highest for the non-zapping class if the maximum smile response is above 0.9. On the contrary for zapping class, majority of the sequences are with the maximum smile response less than 0.1.

For non-zapping class, the probability is the second highest (15.5 percent) when the smile response is less than 0.1. Observations on our data show that a few participants watch the entire ad but display minor smile expression. This means that entertaining content is not the only reason to keep the user engaged. Besides, the interview of participants also shows that a small group of people enjoyed the ad but prefer not to show their feelings through facial expression.

For zapping class, the probability decreases as maximum smile response increases, and reaches the minimum when smile response is between 0.7 and 0.8. However, the probability increases thereafter. After examining the data, we

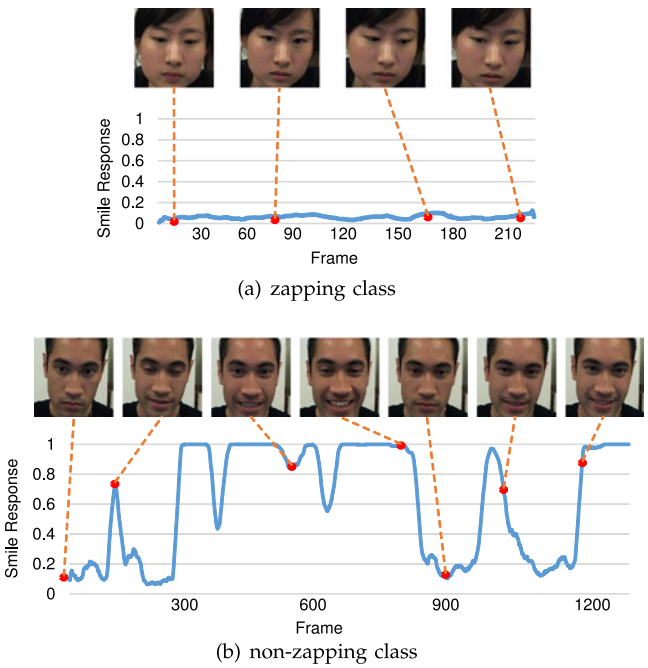


Fig. 7. Sample frames of smile response from zapping and non-zapping classes.

found that several subjects were engaged by the ad and were smiling with high intensity in the beginning. Unfortunately, they zapped right away when the brand's logo or name showed up at the end of the ad. After interviewing with them, we found that most of the people behaved like this because they thought the ad is about to finish. From the advertiser's point of view, this scenario should be considered as a success. However, from the publisher's point of view, they will not get paid since they consider this scenario as zapping [2].

Remark. 1 (Note to Advertisement Publisher). In our analysis, 13.3 percent of the time, users zap because advertiser's brand is displayed at the end when the ad is not finished. In this case, the advertisers take the benefit since users consume the content of the ad. The publisher (such as YouTube), on the other hand, loses revenue since the ad is neither watched for more than 30 seconds nor completed by the user under this circumstance. If the billing policy is changed from "30 seconds" to "27 seconds" or from "complete the entire advertisement" to "complete 90 percent of the advertisement", less zapping is likely to happen in our experiment. In light of this observation, a publisher is suggested to change the billing policy in the aforementioned manner while maintaining the effectiveness of a commercial.

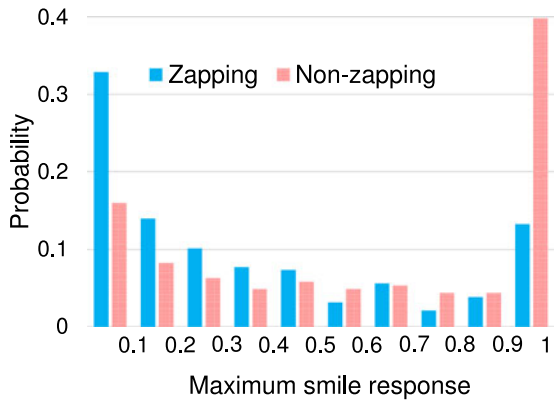


Fig. 8. The distribution of the zapping/non-zapping data based on their maximum smile response.

3.3.4 *The Volume of Smile Response*

In addition to maximum smile response, we also analyze how the volume of the smile response distinguishes the zapping and non-zapping classes. The volume of smile response is defined as the portion of the length of the sequence that is above a certain smile response level. Figs. 9a and 9b are a variation of 2-D cumulative distribution function (CDF) defined as

$$F_{XY}(x, y) = P(X \geq x, Y \geq y), \tag{3}$$

where X is a random variable that measures the portion of an advertisement that is watched, and Y represents smile response. It can be interpreted as follows: if the smile response of a sequence is *above* y for x percent of its entire length, the probability of this event is F.

In both Figs. 9a and 9b, the CDF is 1 when smile response is close to 0 (bottom edge of the figure). This is because all the data satisfy the criteria that the smile response is always above 0. The reason why the upper right corner is of value 0 is because no sequence has a smile response above 0.9 for 90 percent of the time.

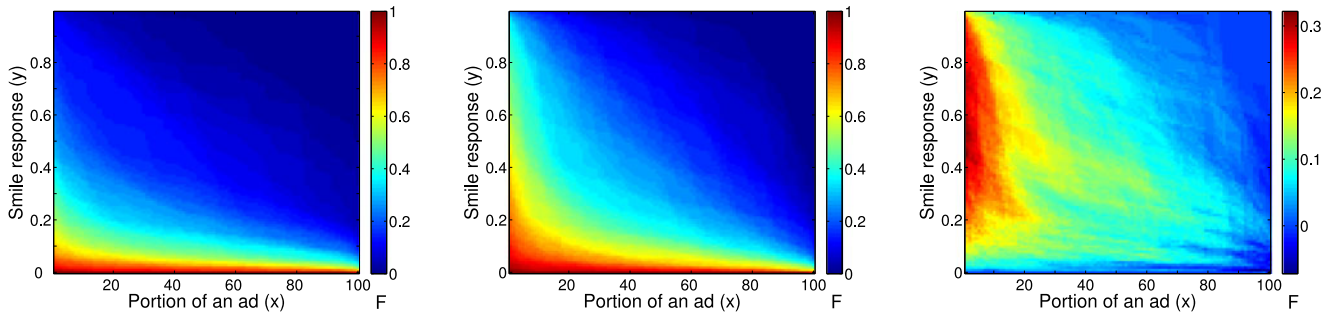


Fig. 9. The analysis of the smile response volume. Figs. 9a and 9b are interpreted as follows: the probability is F if the smile response of a sequence is *above* y for x percent of its entire length. Fig. 9c is generated by subtracting Figs. 9a from 9b, which shows that high smile response is more effective than high volume in distinguishing zapping from non-zapping. (Better viewed in color).

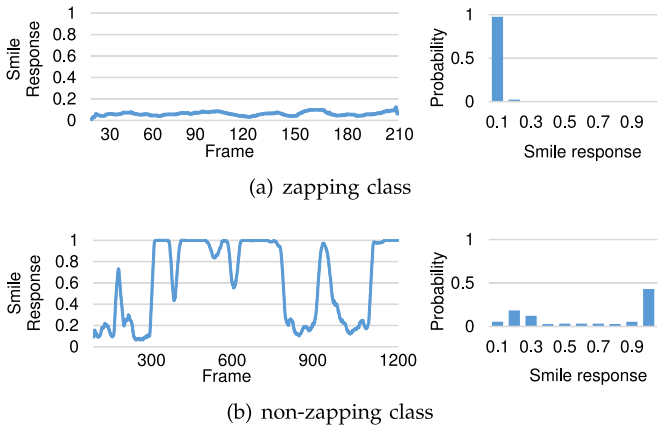


Fig. 10. Smile histogram of zapping and non-zapping sequences. The selected sequences are the same as in Fig. 7.

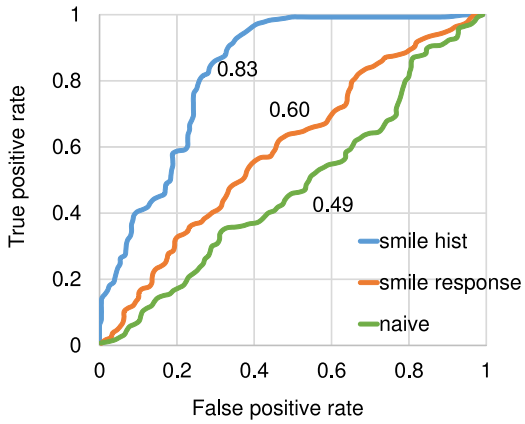


Fig. 11. The ROC plot for zapping/non-zapping classification. The naïve baseline is by assigning test labels based on the class distribution.

Compared to the zapping class, the CDF of the non-zapping class in Fig. 9b is close to symmetrical along the diagonal line from the bottom left corner to the upper right corner. This shows that, for non-zapping class, both the level of the smile response and its volume play important roles in the data distribution. For example, the CDFs are similar for non-zapping class for the following two cases: (1) smile response is above 0.6 for 20 percent amount of the time; (2) smile response is above 0.2 for 60 percent amount of the time.

We show the CDF difference in Fig. 9c by subtracting Fig. 9a from Fig. 9b. The major difference exists where the smile volume is low and the smile response level is high. This informs us that, in distinguishing zapping from non-zapping, high smile response level is more important than high volume.

Remark. 2 (Note to Advertiser). Our statistics shows that users tend to zap less if their smile response level is higher or their smile response is above a certain level for a longer period of time. Moreover, if an advertiser has to choose between "high smile response level + low volume" and "low smile response level + high volume", the former is more effective in preventing zapping. Our experimental result suggests that, practically, it is preferred to design an ad with one or two entertaining scenes that highly impact the user's engagement than to include several entertaining scenes with mediocre impact, if eliciting laugh is the objective of a commercial.

3.4 Zapping Index

3.4.1 The Smile Histogram Feature

Based on the data characteristics, we find that the mean, max, and the volume of smile response of a video sequence are essential for distinguishing zapping from non-zapping. It is natural to use the histogram of the smile response as the key feature.

As shown in Fig. 10, the cumulative smile histogram is calculated for the entire sequence. It is then normalized between 0 and 1. In the two typical examples in Fig. 10, the probability is high when the smile response is low for the zapping class. For non-zapping class, on the contrary, the smile response is more evenly distributed.

3.4.2 Zapping Classification

In order to distinguish zapping from non-zapping sequences, we formulate it as a binary classification problem. The class labels of the data are assigned based on the 90 percent threshold shown in Fig. 5.

During the training phase, the histograms of all the sequences are computed. We use SVM [23] with the radial basis function as the kernel function to train our classifier. The *double-layer* 10-fold cross validation is then carried out to avoid overfitting. The first layer is for parameter optimization and the second layer uses the optimized parameters for model training. The number of bins is then determined to be 10 by the second layer of cross validation. Then a validation set is constructed by randomly generating 4,000 frames from the entire dataset.

In comparison, we provide the baseline result from naïvely assigning labels based on class distribution shown in Fig. 5. In addition, we include the result of using the normalized cumulative smile response of individual sequences. Fig. 11 shows the ROC curve of the aforementioned

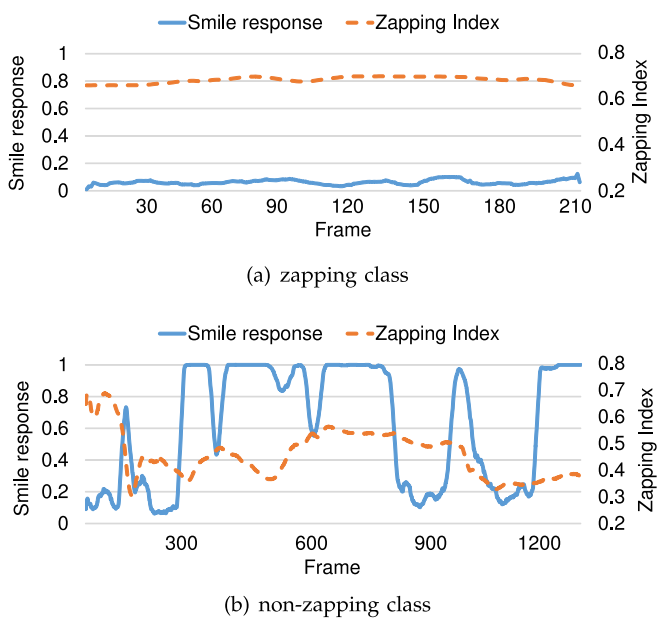
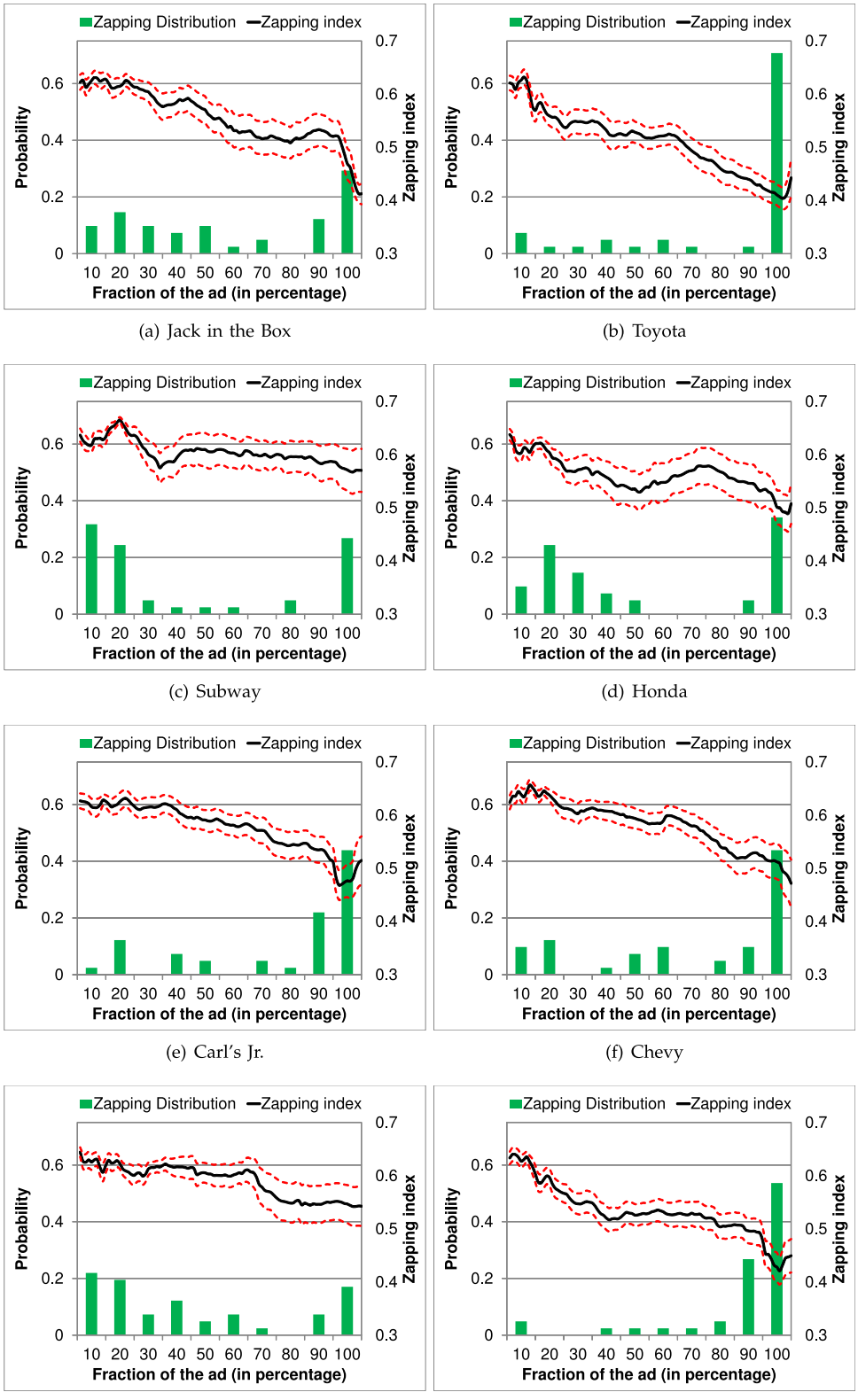


Fig. 12. The moment-to-moment Zapping Index for individual sequences. The selected sequences are the same as in Fig. 7.



(a) Jack in the Box (b) Toyota

(c) Subway (d) Honda

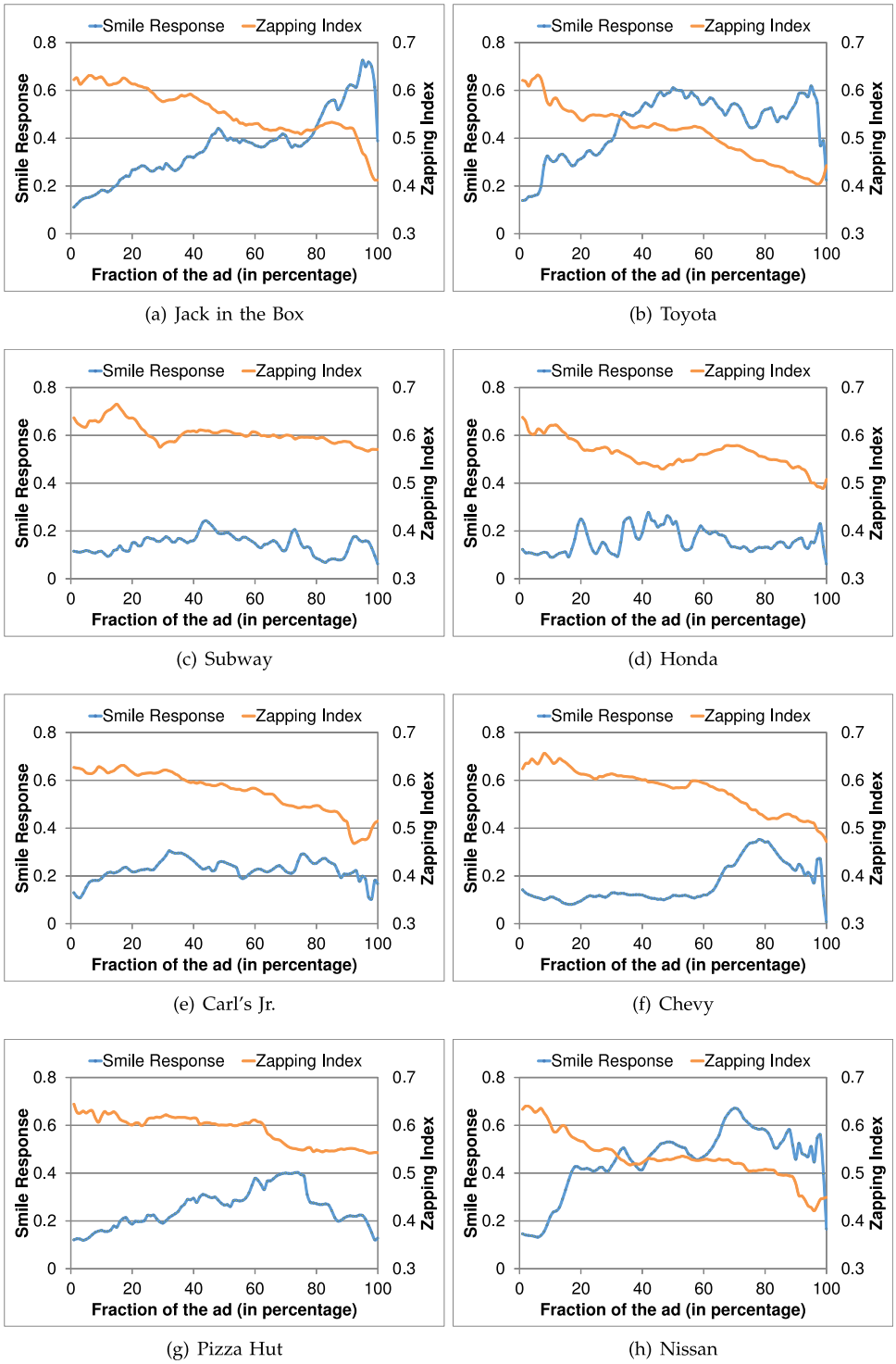
(e) Carl's Jr. (f) Chevy

(g) Pizza Hut (h) Nissan

Fig. 13. The relationship of Zapping Index and zapping distribution for each ad. Zapping distribution is the normalized histogram measured with the primary *y*-axis, while ZI is displayed by the secondary *y*-axis bounded by its standard error (shown with dotted lines). Generally speaking, the more sharply ZI decreases, the less zapping happens.

approaches. As AUC scores illustrated in Fig. 11, smile histogram feature (0.83) significantly outperform smile response feature (0.60), which is congruent with our analysis in Section 3.3 that the mean, maximum, and volume of smile response are essential in characterizing zapping behavior.

During the testing phase, our goal is to measure the moment-to-moment zapping probability. Thus, at each frame, the smile histogram feature is computed, which is then passed to the classifier, and the probability output is considered as the zapping index. The upcoming discussion



(a) Jack in the Box (b) Toyota

(c) Subway (d) Honda

(e) Carl's Jr. (f) Chevy

(g) Pizza Hut (h) Nissan

Fig. 14. The relationship of smile response and Zapping Index for each ad. Broadly speaking, they are in inverse relation. But the ZI is less volatile as compared to smile response.

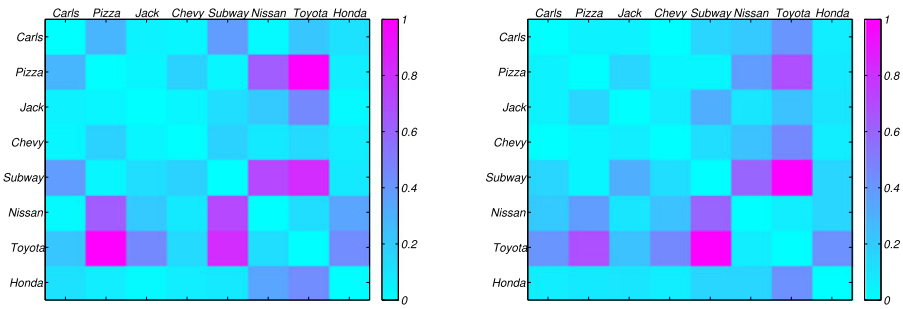
in the next section shows why ZI is a valid measurement for zapping behavior and how it is related to zapping prediction and user preference discovery.

4 Experimental Results

In this section, we explore the characteristics of the Zapping Index on individual sequences, across each ad, and across each ad category. We also visualize their relationships with data distribution that show the popularity of each ad. Moreover, we are able to understand the preference of users to different ad categories.

4.1 Zapping Index on Individual Sequences

According to our design, a larger value of ZI means higher probability of zapping. Fig. 12 provides the Zapping Index for the two running examples. Generally speaking, if a user displays low smile response, the ZI remains around 0.65; the ZI decreases when smile response increases. The ZI value of 0.65 coincides with our discussion related to Fig. 8,



(a) Zapping Distribution (b) Zapping Index

Fig. 15. The pair-wise distance of individual ad in zapping distribution and Zapping Index groups. ZI is highly correlated with viewer's zapping behavior exhibited in zapping distribution, which demonstrates that ZI is a reasonable measure for zapping. (The distances within each group are normalized between 0 and 1.)

when maximum smile response of a sequence is between 0 and 0.1, there is a two thirds chance that it belongs to the zapping class.

As discussed in Section 3.3.3, if a user's smile response reaches the maximum, he or she will most unlikely zap. In Fig. 12b, there might be minor increase of ZI if smile response drops from the maximum. However, the ZI value will remain low even after the increase, which illustrates that the user is less likely to zap.

4.2 Zapping Index vs. Zapping Distribution

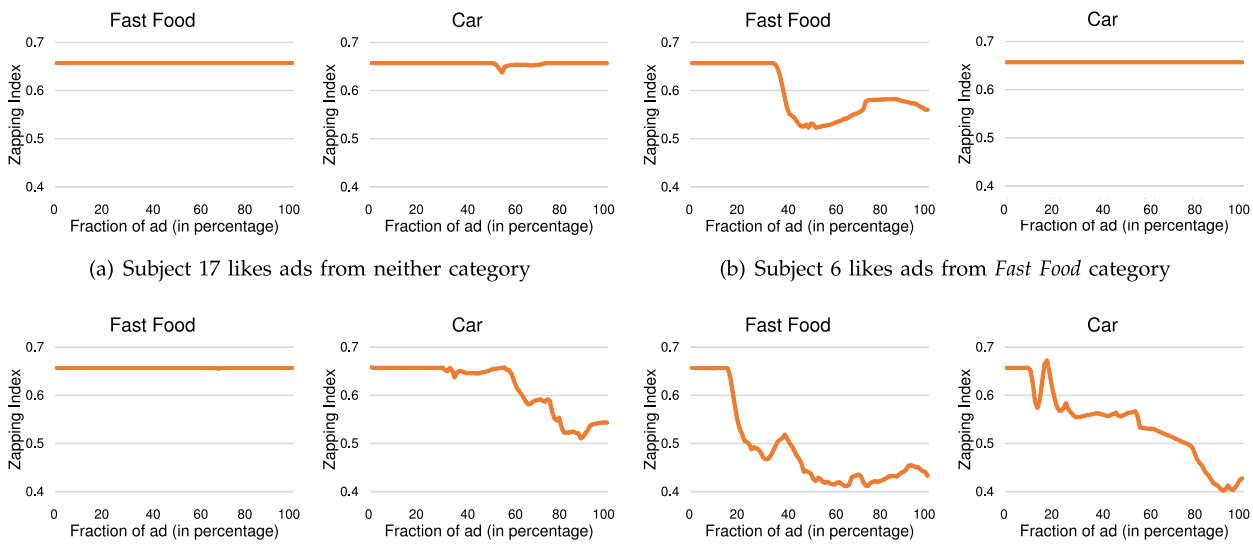
In Fig. 13, we compare the relationship between the average ZI value and the zapping distribution for each ad. We compute the mean and SEM bound of ZI similar to the description in Section 3.3.2.

For ads which have less zapping as shown by the zapping distribution (e.g., Toyota and Nissan), the decrease rate of the moment-to-moment ZI is high which means small likelihood of zapping by users. Indeed, these two ads are the funniest ads among all of our selections. On the contrary, for ads for which a larger number of participants zapped (e.g. Pizza Hut and Subway), the ZI curve only has a slight decrease. Therefore, our ZI measurement correlates with the zapping distribution.

To further demonstrate that ZI is a quality measurement for zapping, we analyze the correlation between ZI and zapping distribution in Fig. 13. We treat a ZI sequence for each ad as a feature and compute the pair-wise Euclidean distance between features. We compute the distance the same way for zapping distribution. The pair-wise distance for both zapping distribution and ZI are plotted in Figs. 15a and 15b, respectively. We observe similar patterns in Figs. 15a and 15b, which means that ZI preserves distance between ads in zapping distribution. For example, the distance is large for *Toyota* and *Pizza Hut* in Fig. 15a; this means that viewers' behavior is dramatically different in watching these two ads. In Fig. 15b, ZI captures the same difference. Therefore, the measurement of ZI is highly correlated with the viewer's zapping behavior.

4.3 Zapping Index vs. Smile Response

We show the comparison of ZI and smile response in Fig. 14. Generally speaking, ZI has an inverse relationship with smile response. One may argue that smile response itself is a good indicator for zapping prediction. However, as we can see from Fig. 14, smile response is a measurement of user's smile expression at every moment, and therefore, it is volatile as time changes. ZI, on the other hand, is a



(a) Subject 17 likes ads from neither category (b) Subject 6 likes ads from *Fast Food* category

(c) Subject 48 likes ads from *Car* category (d) Subject 2 likes ads from both categories

Fig. 16. All four possible types of user preferences represented by ZI. Each ZI pattern is the average of one subject on the entire ad category. Flat ZI response such as both cases in (a) shows no smile response as well as no interest to the ad category, and vice versa.

TABLE 2
Another Advertisement Category: "Running Shoe"

Category	Brand	Ad Name
Running Shoe	Nike	Flyknit Lunar 1+
	Adidas	Boost
	Puma	Mobium and Adaptive
	Under Armour	I Will Innovation

better measurement for predicting zapping. It is less volatile by taking into account the maximum smile and smile occurring volume information overtime (see Fig. 9). Yet, it is also sensitive enough to capture the noticeable changes in smile response.

One interesting pattern that is worth noting in ads such as Jack In The Box, Nissan, Toyota is that there is a major drop for the smile response at the end. After examining the participant's expression as well as the ads themselves, we found out that most participants smile due to the entertaining scene at the end. They tend to smile less as soon as they saw the brand's logo or name. This phenomenon is congruent with our discussion in Section 3.3.3.

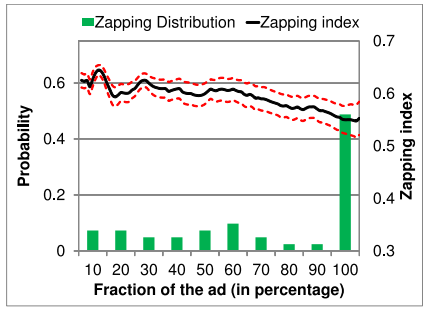
4.4 User Preference from Zapping Index

An important factor, if not the most important one, in considering the advertising campaign solution is the target market [30]. Advertisement is viewed as useful information for the right target, but viewed as harassment for the wrong target. Advertising publishers explore a large data from users to discover the target market. User information such as age, gender, ethnicity, geography, income, lifestyle, online behavior, etc., are leveraged to infer the user preference. The most recent work [31] has demonstrated that smile response is able to reveal whether an ad is liked by a viewer. In this paper, we show that Zapping Index, derived from smile response, is another type of user information which directly shows viewer's preference. Thus, ZI may have a potential impact on the future of advertising. We analyze user preferences for two different ad categories in our experiment. Fig. 16 shows four typical samples of user preferences of two different ad categories expressed by ZI. These four types include: like neither category, like first category but not the second, like the second category but not the first, and like both categories. By classifying the user based on their ZI, it is possible to accurately measure their preferences, which will benefit both the advertising provider and the publisher.

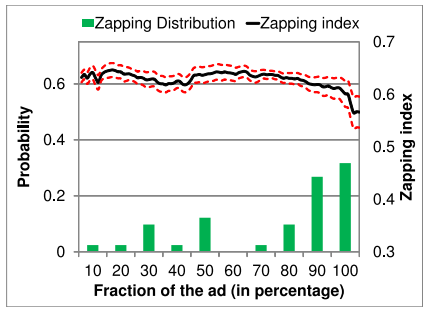
4.5 Limitations

In this work, we only consider the entertaining value of an ad and use smile response to compute ZI for the prediction of a user's zapping probability. However, entertainment is not the only reason that engages a user. For example, information content is another major reason for user engagement [3], in which case, user will not necessarily smile. Under this situation, our ZI measurement may be less effective in predicting zapping probability.

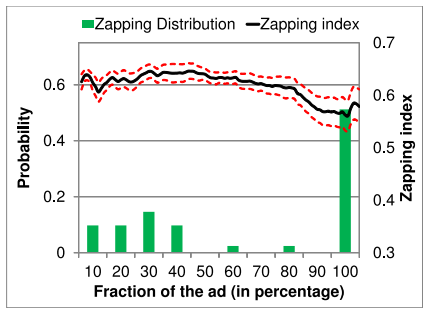
To test our assumption, we selected another four ads from the *running shoe* category, shown in Table 2. We make the selections by following the criteria described in Section 3.2. The only difference is that we selected the high quality ad but not necessarily amusing.



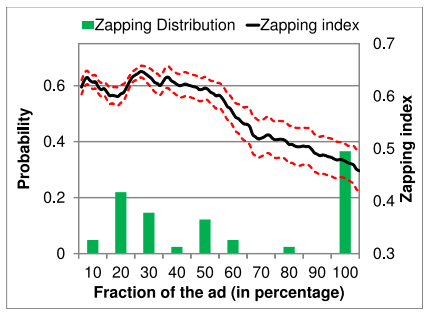
(a) Nike



(b) Adidas



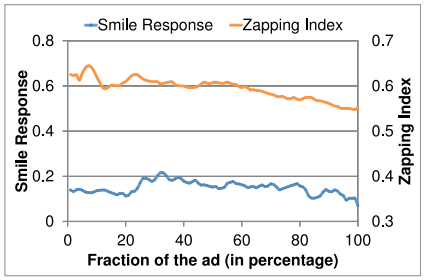
(c) Puma



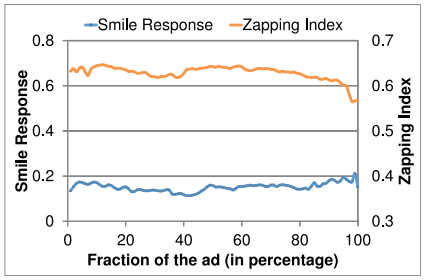
(d) Under Armour

Fig. 17. Zapping Index vs. zapping distribution for the ads in *running shoe* category. ZI is less effective in predicting zapping probability compared to Fig. 13 since soliciting smile is not the intention of these ads. Hence, users tend to zap less even if ZI value is high.

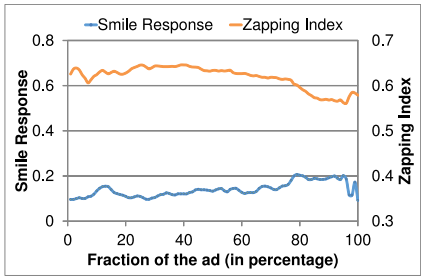
Similar to Figs. 13 and 14, we explore the relationship of ZI vs. zapping distribution and ZI vs. smile response in Figs. 17 and 18, respectively. The ads from the *Running Shoe* category are not intended to make a user laugh but rather to provide information by demonstrating their technology. Hence, the average smile response is low and the ZI does not necessarily decrease in Fig. 18, which shows that users



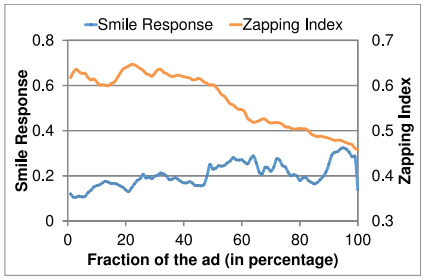
(a) Nike



(b) Adidas



(c) Puma



(d) Under Armour

Fig. 18. Zapping Index vs. smile response for the ad in the *running shoe* category. Smile response is low but this does not always lead to a decrease in the value of ZI.

are likely to zap. However, based on the zapping distribution in Fig. 17, majority of the participants have watched the ads without zapping. Thus, ZI is less effective as a zapping prediction metric when the intention of the ad is not engaging people through entertaining factors.

5 Conclusions

This paper explored the automated facial expression recognition in the application of online advertising. We demonstrated that users' zapping behavior has a close relationship with their smile response. We created an advertising evaluation metric, Zapping Index, to measure a user's zapping probability. A higher value of ZI reveals that the user has a higher chance of zapping. ZI can also be used to measure a user's preference to different categories of commercials. This is beneficial to advertisers as well as to ad publishers. In the future, it would be interesting to analyze the effects of other facial expressions or other representations of facial expression (e.g., Action Unit) on zapping behavior. Moreover, as demonstrated in [31], dynamic features of expression outperform static features in user preference prediction. Thus, incorporating dynamic features in ZI classification is also a promising direction for future work.

Acknowledgments

This work was supported in part by NSF grants 0905671 and 0727129. The contents of the information do not reflect the position or policy of the U.S. Government.

References

[1] E. Gabarron, L. Fernandez-Luque, M. Armayones, and A. Y. Lau, "Identifying measures used for assessing quality of Youtube videos with patient health information: A review of current literature," *Int. J. Med. Res.*, vol. 2, no. 1, p. e6, 2013.

[2] M. Pashkevich, S. Dorai-Raj, M. Kellar, and D. Zigmond, "Empowering online advertisements by empowering viewers with the right to choose," *J. Advertising Res.*, vol. 52, pp. 65–71, 2012.

[3] 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.*, vol. 4, pp. 437–453, 2003.

[4] K. Poels and S. Dewitte, "How to capture the heart? Reviewing 20 years of emotion measurement in advertising," Dept. Commun. Studies, Katholieke Universiteit Leuven, Leuven, Belgium, Tech. Rep. MO-0605, 2006.

[5] P. Ekman, "Facial expression and emotion," *Amer. Psychol.*, vol. 48, pp. 384–392, 1993.

[6] T. Teixeira, M. Wedel, and R. Pieters, "Emotion-induced engagement in internet video advertisements," *J. Marketing Res.*, vol. 49, pp. 144–159, 2012.

[7] D. McDuff, R. Kaliouby, and R. Picard, "Crowdsourcing facial responses to online videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 4, pp. 456–468, Fourth Quarter 2012.

[8] P. Hyde, E. Landry, and A. Tipping, "Making the perfect marketer," *Strategy + Bus.*, vol. 37, pp. 37–43, 2004.

[9] C. Shan, "Smile detection by boosting pixel differences," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 431–436, Jan. 2012.

[10] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Toward practical smile detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2106–2111, Nov. 2009.

[11] C. Hu, Y. Chang, R. Feris, and M. Turk, "Manifold based analysis of facial expression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2004, p. 81.

[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 *Proc. Int. Conf. Autom. Face Gesture Recognit.*, 2006, pp. 155–160.

[13] M. Valstar, I. Patras, and M. Pantic, "Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2005, p. 75.

[14] I. Kotsia, and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 172–187, Jan. 2007.

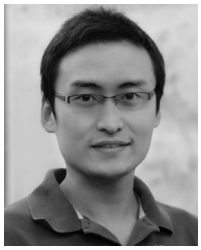
[15] G. Zhao, and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[16] S. Yang and B. Bhanu, "Understanding discrete facial expressions in video using an emotion avatar image," *IEEE Trans. Syst., Man, Cybern., Part B*, vol. 42, no. 4, pp. 980–992, Aug. 2012.

[17] P. Gustafson, and S. Siddarth, "Describing the dynamics of attention to tv commercials: A hierarchical Bayes analysis of the time to zap an ad," *J. Appl. Stat.*, vol. 34, pp. 585–609, 2007.
[18] R. Kooij, K. Ahmed, and K. Brunnström, "Perceived quality of channel zapping," in *Proc. 5th IAESTED Int. Conf. Commun. Syst. Netw.*, 2006, pp. 156–159.
[19] P. Siebert, T. Van Caenegem, and M. Wagner, "Analysis and improvements of zapping times in IPTV systems," *IEEE Trans. Broadcasting*, vol. 55, no. 2, pp. 407–418, Jun. 2009.
[20] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
[21] S. Yang, L. An, B. Bhanu, and N. Thakoor, "Improving action units recognition using dense flow-based face registration in video," in *Proc. Int. Conf. Autom. Face Gesture Recognit.*, 2013, pp. 1–8.
[22] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Proc. 3rd Int. Conf. Image Signal Process.*, 2008, pp. 236–243.
[23] C.-C. Chang and C.-J. Lin. (2001). LIBSVM: A Library for Support Vector Machines [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm
[24] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, no. 6, pp. 902–913, 2010.
[25] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.
[26] 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., Part A*, vol. 38, no. 1, pp. 149–161, Jan. 2008.
[27] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2010, pp. 94–101.
[28] L. An, S. Yang, and B. Bhanu, "Efficient smile detection by extreme learning machine," *Neurocomputing*, 2014, http://dx.doi.org/10.1016/j.neucom.2014.04.072.
[29] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2013, pp. 881–888.
[30] P. Kotler, *Marketing Management*. Noida, UP, India: Pearson Education India, 2009.
[31] D. McDuff, R. Kaliouby, T. Senechal, D. Demirdjian, and R. Picard, "Automatic measurement of ad preferences from facial responses gathered over the internet," *Image Vis. Comput.*, vol. 32, no. 10, pp. 630–640, 2014.

Songfan Yang (S'10-M'14) received the BS degree in electrical engineering from Sichuan University, Chengdu, China, in 2009, and the MS and PhD degrees in electrical engineering from the University of California, Riverside, in 2011 and 2014, respectively. He has joined Sichuan University as an associate professor. His current research interests include affective computing and human behavior understanding. He holds the Best Entry Award of the FG 2011 Facial Expression Recognition and Analysis Emotion challenge (FERA) competition. He is a member of the IEEE.

Mehran Kafai (S'11-M'13) received the MSc degree in computer engineering from Sharif University of Technology, Tehran, Iran, in 2005, the MSc degree in computer science from San Francisco State University in 2009, and the PhD degree in computer science from the Center for Research in Intelligent Systems (CRIS), University of California, Riverside, in 2013. He is a research scientist at Hewlett Packard Laboratories in Palo Alto, California. His recent research has been concerned with secure computation, information retrieval, and big-data analysis. He is a member of the IEEE.



Le An received the BEng degree in telecommunications engineering from Zhejiang University, Hangzhou, China, in 2006, and the MS degree in electrical engineering from Eindhoven University of Technology, Eindhoven, The Netherlands, in 2008. He is currently working toward the PhD degree in electrical engineering at the Center for Research in Intelligent Systems at the University of California, Riverside. His research interests include image processing, computer vision, pattern recognition, and machine learning. He is a student member of the IEEE.

Bir Bhanu (S'72-M'82-SM'87-F'95) received the SM and EE degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, the PhD degree in electrical engineering from the Image Processing Institute, University of Southern California and the MBA degree from the University of California, Irvine. He is the distinguished professor of electrical and computer engineering, interim chair of Bioengineering Department and cooperative professor of computer science, mechanical engineering, and the director of the Center for Research in Intelligent Systems (CRIS) and the Visualization and Intelligent Systems Laboratory (VISLab) at the University of California, Riverside (UCR). In addition, he serves as the director of NSF IGERT on Video Bioinformatics at UCR. He has been the principal investigator of various programs for NSF, DARPA, NASA, AFOSR, ONR, ARO, and other agencies and industries in the areas of video networks, video understanding, video bioinformatics, learning and vision, image understanding, pattern recognition, target recognition, biometrics, autonomous navigation, image databases, and machine-vision applications. He has published seven authored and three edited books. He is the holder of 18 (four pending) patents. He has published more than 500 reviewed technical publications, including more than 125 journal papers and 45 book chapters. He has received University and Industry Awards for research excellence, outstanding contributions and team efforts, including graduate advisor/mentor award of the university. He has also received many outstanding journal and best conference awards. He is a fellow of the IEEE, AAAS, IAPR, and SPIE. He served on the IEEE Fellow Committee from 2010-12.