You Are Known by How You Vlog: Personality Impressions and Nonverbal Behavior in YouTube

Joan-Isaac Biel[1,2] and **Oya Aran**[1] and **Daniel Gatica-Perez**[1,2]
[1]Idiap Research Institute
[2]Ecole Polytechnique Fédérale de Lausanne (EPFL)
Switzerland
{jibiel,oaran,gatica}@idiap.ch

Abstract

An increasing interest in understanding human perception in social media has led to the study of the processes of personality self-presentation and impression formation based on user profiles and text blogs. However, despite the popularity of online video, we do not know of any attempt to study personality impressions that go beyond the use of text and still photos. In this paper, we analyze one facet of YouTube as a repository of brief behavioral slices in the form of personal conversational vlogs, which are a unique medium for self-presentation and interpersonal perception. We investigate the use of nonverbal cues as descriptors of vloggers' behavior and find significant associations between automatically extracted nonverbal cues for several personality judgments. As one notable result, audio and visual cues together can be used to predict 34% of the variance of the Extraversion trait of the Big Five model. In addition, we explore the associations between vloggers' personality scores and the level of social attention that their videos received in YouTube. Our study is conducted on a dataset of 442 YouTube vlogs and 2,210 annotations collected using Amazon's Mechanical Turk.

Introduction

People tend to use social media to express and communicate their personality, and the content and behavior they display, explicitly or not, convey information accurately perceived by others (Back et al. 2010). As the amount of users participating in social media outlets and the content available on these increase, there is a need to understand how the processes of self-personality presentation and personality impression formation take place. This is evidenced by the great interest of the social media research community on these topics (Back et al. 2010; Evans, Gosling, and Carroll 2008; Yarkoni 2010).

Nonverbal behavior is an effective way of expressing aspects of identity such as age, occupation, culture, and personality, and it is consequently used to make inferences about them (Ambady and Rosenthal 1992). Nonverbal cues have been shown to be useful to characterize social constructs related to conversational interaction in both social psychology (Knapp and Hall 2005) and social computing (Pentland 2008; Gatica-Perez 2009). In social media, a recent study on YouTube suggested that, in addition to the content, nonverbal behavior plays a role in conversational vlogs, in ways that might bear similarities with face to face interactions (Biel and Gatica-Perez 2010a). In particular, there is initial evidence that some audio, visual, and multimodal nonverbal cues extracted from conversational vlogs are significantly correlated with the level of attention that videos received. Clearly, conversational vlogs are a unique medium for self-presentation and interpersonal perception in social media, going beyond the use of text and still photos, which may partly explain the popularity of this format among online video users. However, despite the 35 hours of video uploaded per minute (and growing) reported by YouTube in their official blog (Nov. 2010), online video has received little attention from the social media community (Biel and Gatica-Perez 2010b).

In this paper, we address the study of personality impressions in vlogging, under the lens of audiovisual behavioral analysis. Specifically, we investigate the use of nonverbal behavioral cues as descriptors of vloggers' behavior and their association to the process of impression formation in this type of social media. This is relevant because, to our knowledge, the existing attempts to study personality impressions in social media have focused on text and still photos from social network users' profiles (Evans, Gosling, and Carroll 2008) and blogs (Yarkoni 2010). In addition, nonverbal behavior conveys information that is generally difficult to control, which might differ from most of the content that people post on their user profiles and blogs.

Our study consists of three main contributions. First, we examine the reliability of personality judgments from online video watching in a sample of 442 vloggers and 2,210 annotations obtained through crowdsourcing using Amazon's Mechanical Turk. Second, we investigate the links between automatically extracted nonverbal cues from audio and video and personality judgments, in addition to the suitability of these descriptors for personality prediction. Finally, we explore the relation between the personality scores' distribution of vloggers and the levels of attention received by their vlogs, as a first attempt to study whether different personality traits could be connected to differences in outcome measures of social media participation.

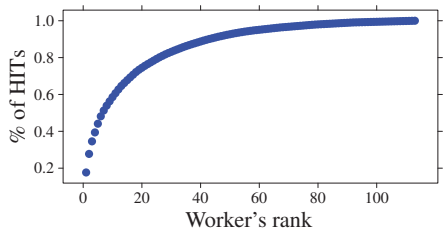

Figure 1: Cumulative frequency distribution of MTurk annotations (in percentages). Workers are ranked based on the number of HITs completed (top ranked worker completed 17% of the HITs).

Methodology

Vlog Collection

We collected videos from YouTube with a keyword-based search for "vlog" and "vlogging" using the API, and we manually filtered the retrieved results to gather a sample of conversational vlogs featuring one person only. We then restricted the sample to one video per vlogger, resulting in a final dataset of 442 vlogs of which 47% (208) corresponded to male and 53% (234) to female vloggers. We limited the size of the dataset in order to bound the amount of annotations required for our experiments.

We automatically processed videos to obtain the first conversational minute of each vlog (the specifics of this procedure are not presented here for space reasons). Using "thin slices" has proven a suitable approach to study a wide range of constructs, including personality traits, affective states, status or dominance, etc. (Ambady and Rosenthal 1992). In fact, for the specific study of personality judgments, some research suggested that a few seconds are enough to make accurate judgments (Carney, Colvin, and Hall 2007).

Personality Annotations

We used Amazon's Mechanical Turk (MTurk) crowdsourcing platform in order to obtain zero-acquaintance judgments of personality. In each Human Intelligence Task (HIT) we asked workers to watch one vlog, and to answer the TIPI questionnaire, a 10-item measure of the Big Five personality dimensions (Gosling, Rentfrow, and Swann 2003), about the person appearing in the vlog. We specifically asked the workers to watch the totality of the one minute vlog, and we disabled the HTML questionnaire until the video had reached the end. In addition to logging the working time reported by MTurk, we also controlled for real time of video watching (to detect if any workers were playing forward the video), and the time spent on the questionnaire.

In total, we posted 442 different HITs to be completed five times each (2,210 HITs in total), and we restricted them to US and Indian workers with HIT acceptance rates of 95% or higher. As shown in Figure 1, HITs where completed by 113 different annotators with a substantial variation on their contribution in number of HITs. The average time of questionnaire completion (i.e., not including the video) was 36.1s, compared to the one minute suggested by Gosling, Rentfrow, and Swann (2003). This result agrees with other recent studies in MTurk where completion times of annotations were reduced with respect to experts' working time, which can be justified by the economic motive of MTurk workers (Soleymani and Larson 2010).

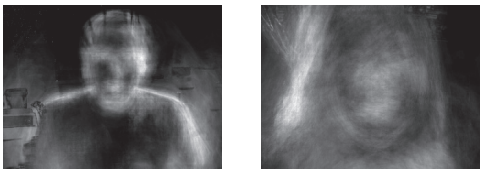

Figure 2: wMEI Images for two vlogs.

Feature Extraction

We automatically extracted nonverbal cues from both audio and video with the purpose of characterizing vloggers' behavior. Given the conversational nature of vlogs, in our study, we focus on the use of nonverbal cues that have shown to be effective in the study of conversational interactions in psychology (Knapp and Hall 2005) and social computing (Pentland 2008; Gatica-Perez 2009). These cues have the advantages of being relatively easy to compute and robust.

Audio Features. We automatically extracted a set of audio cues using the toolbox developed by the Human Dynamics group at MIT Media Lab (Pentland 2008), computed on the one minute vlog slices. This toolbox implements a two-level hidden Markov model (HMM) to segment the audio in voiced/unvoiced and speech/non-speech regions. These segmentations are then used to extract various statistics on speaking activity (speaking time, speaking turns, and voicing rate) as well as emphasis patterns (energy, pitch, and autocorrelation peaks). These nonverbal cues measure how people speak (how much, how loud, how fast, etc) rather than what they say (Biel and Gatica-Perez 2010b).

Visual Features. We automatically extracted a set of visual cues as descriptors of the overall visual activity of the vlogger throughout the video. In this paper, we propose a modified version of motion energy images (Bobick and Davis 2001), that we call "Weighted Motion Energy Images" (wMEI). The wMEI is calculated as:

$$wMEI = \sum_{t=0}^{T} (D_t), \tag{1}$$

where D_t is a binary image that shows the moving pixels in frame t, and T is the duration of the vlog in frames. A wMEI is normalized by dividing all the pixel values by the maximum pixel value. Thus, a normalized wMEI contains the accumulated motion through the video as a gray-scale image, where each pixel's intensity indicates the visual activity in the pixel (brighter pixels correspond to regions with higher motion). From the normalized wMEIs, we extract simple statistical features as descriptors of the vlogger's body activity such as the entropy, mean, median, and center of mass (in horizontal, and vertical dimensions). To compensate for different video sizes, all images are previously resized to 320x240. Figure 2 shows two examples of wMEI images.

Analysis and Results

We divide our analysis in three parts. First, we examine the reliability of MTurk annotations. Second, we investigate the association between nonverbal behavior and personality. Finally, we explore the association between personality traits and social attention.

Personaliy Trait	*ICC(1,1)*	*ICC(1,k)*
Extraversion	$.40^{***}$	$.77^{***}$
Agreableness	$.27^{***}$	$.65^{***}$
Conscientiousness	$.14^{***}$	$.45^{***}$
Emotional Stability	$.13^{***}$	$.42^{***}$
Openness to Exp.	$.15^{**}$	$.47^{***}$

Table 1: Intra-class correlation coefficients for the Big Five personality traits $^{**}p < .001$,$^{***}p < .0001$.

MTurk Annotations' Quality

Though behavioral research using MTurk has reported a fair work quality in a wide range of tasks, it is not entirely clear how this extends to multimedia content annotations solely based on observers' personal judgments (Soleymani and Larson 2010). To explore the reliability of our annotations, we computed two intra-class correlation (ICC) measures that are adequated for our annotation task, where each target is annotated by k judges randomly selected from a population of K judges, $k < K$ (Shrout and Fleiss 1979).

Table 1 shows the ICCs for each personality trait. The ICC(1,1) gives a reliability measure of the single MTurk annotations. In turn, the ICC(1,k) provides a reliability measure for the average of the five MTurk annotations available for each target, which we use as the aggregated personality score for each vlog in the next two sections. Both measures show significant ($p < 10^{-3}$) reliabilities for all the personality traits. In particular, ICC(1,1) shows moderate to low reliabilities for the single MTurk annotations (.15 < ICC(1,1) < .40), whereas the ICC(1,k) display moderate reliabilities for the aggregated annotations (.47 < ICC(1,k) < .77). Interestingly, Extraversion estimates show the highest reliability in both cases, which agrees with findings reported in other studies (Back et al. 2010). Unfortunately, these reliabilities cannot be directly compared to most social media and psychology studies on personality, for two reasons. First, common measures of inter-rater reliability assume annotators to rate the full target set, a requirement easily violated in crowdsourcing settings unless tasks are specifically designed with that intent (Soleymani and Larson 2010), which can be infeasible for large datasets. Second, compared to other measures, ICC(1,1) and ICC(1,k) do not account for individual variance introduced by specific annotators.

Nonverbal Behavior and Personality

We study the individual association between nonverbal behavior and personality judgments based on the pairwise Pearson's correlations of nonverbal cues and personality traits for the 442 vlogs. For each vlog, the personality scores are obtained from the average ratings given by the five MTurk workers that completed the corresponding HIT.

Table 2 shows the correlations between nonverbal cues and the Big Five. Note that 15 out of the 20 nonverbal cues show significant correlations with at least one personality trait, which in most cases correspond to traits with higher reliabilities, a tendency reported elsewhere (Gosling, Ko, and Mannarelli 2002). Extraversion (E) shows the largest number of significant correlations, followed by Openness to experience (OE), and Agreeableness (A).

Audio cues	E	A	C	ES	OE
Speaking time	$.18^{***}$	.01	$.25^{***}$	.07	$.10^{\dagger}$
Voice rate	.04	$.10^{\dagger}$	$.10^{*}$	.07	.05
# Speech turns	$-.10^{\dagger}$	.03	$-.02$	.02	$-.05$
F0 (m)	$.16^{**}$	.08	$-.09$	$-.05$	.05
F0 (s)	$-.14^{*}$	$-.14^{**}$	.01	$-.00$	$-.04$
F0 conf. (m)	$.23^{***}$	$.12^{\dagger}$	.03	.03	.07
F0 conf. (s)	$.17^{**}$	.11	.02	.04	.09
Loc R0 pks (m)	$.22^{**}$	.02	$-.04$	$-.03$	.06
Loc R0 pks (s)	$-.10^{\dagger}$	$-.10^{\dagger}$	$-.09$	$-.06$	$-.08$
# R0 pks (m)	$.15^{*}$	$-.03$	$-.04$	$-.01$	$-.01$
# R0 pks (s)	.05	$-.03$	$-.08$	$-.02$	$-.06$
Energy (m)	$.15^{*}$	$-.05$	$-.03$	$-.03$	.03
Energy (s)	.02	$-.05$	$-.06$	$-.08$	$-.00$
D Energy (m)	$.24^{***}$	$-.09$	$-.06$	$-.08$	$.10^{\dagger}$
D Energy (s)	$-.01$	$-.01$	$-.03$	.02	$-.06$
Visual cues	E	A	C	ES	OE
wMEI (e)	$.37^{***}$	.01	$-.12^{**}$	$-.02$	$.22^{***}$
wMEI (m)	$.31^{***}$	.04	$-.09$	.03	$.25^{***}$
wMEI (md)	$.28^{***}$	.05	$-.10^{\dagger}$	.02	$.23^{***}$
wMEI H Com	.05	$-.05$	$-.01$	.04	$-.01$
wMEI V Com	$-.01$	$-.04$	$-.06$	$-.05$	$-.05$

Table 2: Pearson's correlation coefficients between nonverbal cues and the Big Five personality traits ($^{\dagger}p < .05$,$^{*}p < .01$,$^{**}p < .001$,$^{***}p < .0001$, m = mean, md = median, s = mean-scaled standard deviation, e = entropy, com = center of mass, H = horizontal, V = vertical).

Among audio features, speaking time shows significant correlations with Conscientiousness (C), E, and OE. This is relevant because speaking activity patterns have consistently shown significant effects on the prediction of several social constructs in multiple conversational scenarios (Knapp and Hall 2005). Other correlations also are backed up with findings in psychology. For example, E appears to be negatively correlated with the number of speech turns, which agrees with findings that associate extraverts with higher fluency (Knapp and Hall 2005). Compared to speaking patterns, the statistics (mean, and mean-scaled standard deviations) of voice quality measures such as energy, delta energy (D Energy), pitch (F0), pitch confidence (F0 conf), and autocorrelation peaks (location, Loc R0 pks; and number, # R0 pks) show higher values of correlation with E, and with few exceptions, they do not show correlations with the rest of the Big Five. In addition, the correlations for mean values of energy and F0 are related to findings that associate extraverted to talking louder and with higher pitch, whereas negative mean-scale standard deviations are associated with having higher vocal control (Knapp and Hall 2005).

Our proposed visual descriptors of activity are, among all features, the ones showing the highest correlation values, doing so with E, as well as with OE and C. Visual activity (as measured by the entropy, mean, and median of wMEI features) is positively correlated with E and OE, and negatively correlated with C. Although relatively few works have automatically extracted visual activity cues related to body motion, and focus of attention for nonverbal behavior (Jayagopi et al. 2009), higher levels of activity are typically associated to enthusiastic, energetic, and dominant people.

Measure	E	A	C	ES	OE
Social Attention	$.93^{***}$	$-.05$	$.62^{*}$	.37	$.78^{***}$

Table 3: Pearson's correlation between vloggers' personality traits and social attention in YouTube ($^{*}p < .01, ^{***} p < .0001$).

We used a step-wise linear regression procedure to measure the power of nonverbal cues to predict the personality scores. Unsurprisingly, we only observed significant results for the E trait, for which we could predict 24% and 14% of the variance using speech ($F = 12.47, p < 10^{-3}$) and visual features ($F = 15.25, p < 10^{-3}$), respectively. Finally, combining features from both modalities, the model predicted a total of 34% of the variance ($F = 13.66$, $p < 10^{-3}$).

Vloggers' Personality and Social Attention

We study the association between vlogger's personality and social media attention in YouTube. We define the *average level of attention* $\hat{v}$ of a set of N videos as the median number of their views v_n, $\hat{v} = median\{\log v_n\}_{n=1}^{N}$ (Biel and Gatica-Perez 2010a). For each personality trait t, we divide vloggers into (roughly equally-sized) L groups corresponding to L personality score levels. Then, we compute Pearson's correlation between the average personality scores $\hat{s_i(t)} = mean\{s_n\}_{n=1}^{N_i}$ and the average level of attention $\hat{v_i}$ obtained by each group $i = 1 \dots L$.

Table 3 shows the correlations between the Big Five score levels and social attention in YouTube, for $L = 20$. More Extraverted, Open to experience, and Conscientious vloggers are associated to higher average levels of attention in YouTube, as indicated by the positive values of the correlations (the three personality traits correspond to the most reliable personality judgments as discussed previously). Intuitively, it is reasonable to think that vloggers scoring higher in personality traits such as E or OE, may result more appealing or interesting to watch, because of the way they create their videos. In addition, these type of personalities are more likely to be active, socially involved, and recognized in the vlogger community.

Conclusions

We presented what, to our knowledge, is the first study on personality impressions from brief behavioral slices of online videos extracted from YouTube. We examined the use of automatically extracted nonverbal cues from audio and video as descriptors of vloggers' behavior, and found that some of them are correlated with mean personality judgments for most of the Big Five traits. In addition, we show that the combination of audio and visual cues can be used to predict 34% of the variance for the Extraversion trait, which corresponds to the most reliably judged personality trait. Furthermore, we found that videos of vloggers scoring higher on the Extraversion, Openness to experience, and Conscientiousness are positively associated to higher average levels of attention in YouTube, which may show how these vloggers are perceived by people and how they interact with the social media community. Clearly, no causality effects are implied.

The reliability measures provided suggest that crowdsourcing may be useful to collect annotations from conversational vlogs. However, we acknowledge that averaging multiple MTurk annotations is a rather "naive" aggregation method, which might, in turn, limit the effects measured in our analysis. Future work can take several directions. We intend to examine alternative ways of aggregating multiple annotations to obtain more reliable personality judgments. We also plan to explore the suitability of automatically extracted nonverbal cues for vlogger classification based on personality. Finally, we would be interested in comparing the accuracy of crowdsourced personality judgments and self-reported personality, which has not yet been studied in vlogs.

Acknowledgments We thank the support of the Swiss National Center of Competence (NCCR) on Interactive Multimodal Information Management (IM)2 and the EU FP7 MC IEF project Automatic Analysis of Group Conversations via Visual Cues in Nonverbal Communication (NOVICOM).

References

Ambady, N., and Rosenthal, R. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences : a meta-analysis. *Psychol Bull* 111(2):256–274.

Back, M. D.; Stopfer, J. M.; Vazire, S.; Gaddis, S.; Schmukle, S. C.; Egloff, B.; and Gosling, S. D. 2010. Facebook profiles reflect actual personality not self-idealization. *Psychol Sci* (21):372–374.

Biel, J.-I., and Gatica-Perez, D. 2010a. Vlogcast yourself: Nonverbal behavior and attention in social media. In *Proc. of ICMI-MLMI*.

Biel, J.-I., and Gatica-Perez, D. 2010b. Voices of vlogging. In *Proc. of AAAI ICWSM*.

Bobick, A. F., and Davis, J. W. 2001. The recognition of human movement using temporal templates. *IEEE T Pattern Anal* 23.

Carney, D. R.; Colvin, C. R.; and Hall, J. A. 2007. A thin slice perspective on the accuracy of first impressions. *J Res Pers* 41(5):1054–1072.

Evans, D. C.; Gosling, S. D.; and Carroll, A. 2008. What elements of an online social networking profile predict target-rater agreement in personality impressions. In *Proc. of AAAI ICWSM*.

Gatica-Perez, D. 2009. Automatic nonverbal analysis of social interaction in small groups: A review. *Image Vision Comput* 27.

Gosling, S.; Ko, S.; and Mannarelli, T. 2002. A room with a cue: personality judgments based on offices and bedrooms. *J Pers Soc Psychol* 82:379–98.

Gosling, S. D.; Rentfrow, P. J.; and Swann, W. B. 2003. A very brief measure of the big five personality domains. *J Res Pers* 37:504–528.

Jayagopi, D. B.; Hung, H.; Yeo, C.; and Gatica-Perez, D. 2009. Modeling dominance in group conversations using nonverbal activity cues. *Trans Audio Speech Lang Proc* 17(3).

Knapp, M. L., and Hall, J. 2005. *Nonverbal communication in human interaction*. New York: Holt, Rinehart and Winston.

Pentland, A. S. 2008. *Honest Signals: How They Shape Our World*, volume 1 of *MIT Press Books*. The MIT Press.

Shrout, P., and Fleiss, J. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 86(2):420–428.

Soleymani, M., and Larson, M. 2010. Crowdsourcing for affective annotation of video: Development of a viewer-reported boredom corpus. In *Proc. of the SIGIR 2010 (CSE Workshop)*.

Yarkoni, T. 2010. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *J Pers Soc Psychol* 44:363–373.