

Persuasiveness in Social Multimedia: The Role of Communication Modality and the Challenge of Crowdsourcing Annotations

Sunghyun Park
University of Southern California
12015 Waterfront Dr.
Playa Vista, CA 90094
park@ict.usc.edu

ABSTRACT

With an exponential growth in social multimedia contents online, there is an increasing importance of understanding why and how some contents are perceived as persuasive while others are ignored. This paper outlines my research goals in understanding human perception of persuasiveness in social multimedia contents, which involve studying how different communication modalities influence our perception and identifying key verbal and nonverbal behaviors that eventually lead us to believe someone is convincing and influential. For any research involving in-depth human behavior analysis, it is imperative to obtain accurate annotations of human behaviors at the micro-level. In addition to investigating persuasiveness, this work will also provide to the research community convenient web-based annotation tools, effective procedures for obtaining high-quality annotations with crowdsourcing, and evaluation metrics to fairly and accurately measure the quality and agreement of micro-level behavior annotations.

Categories and Subject Descriptors

J.4 [Computer Applications]: Social and Behavioral Sciences – *psychology, sociology*.

H.5.2. [Information Interfaces and Presentation]: User Interfaces – *graphical user interface, evaluation/methodology*.

General Terms

Algorithms, Measurement, Performance, Experimentation, Human Factors.

Keywords

Persuasiveness; perception; multimodal; crowdsourcing; OCTAB; behavior annotations

1. INTRODUCTION

With the advent of the Internet age and social network services, social multimedia contents are becoming an ever more important source of communication and exchange of ideas. More than 100

hours of videos are being posted online on YouTube every minute¹, and people on Facebook and Twitter constantly share their opinions on everything from new commercial products to controversial political events. Due to the unprecedented scale and speed of such new form of communication, social multimedia contents are exerting more and more influence in shaping people's opinions. This societal phenomenon naturally gives importance to understanding factors related to persuasiveness in social multimedia contents.

Persuasive communication, defined as “*any message that is intended to shape, reinforce, or change the responses of another, or others*” [12], plays an important role in creating social influence and changing others' attitudes [19, 26]. Studies on persuasion have identified a number of social and behavioral cues associated with persuasiveness including verbal and nonverbal behaviors, physical appearance, psychological attributes, social dynamics, and social roles [8, 21].

In order to perform in-depth analysis of persuasiveness in social multimedia contents, it is imperative to make accurate annotations of human behaviors at the micro-level such that the precise start and end of an event or a behavior are identified. Unfortunately, automatic methods of annotating human behaviors are still far from adequate, and researchers often have no choice but to rely on manual annotations that are costly both in terms of time and budget. With a recently emerging paradigm of crowdsourcing, such manual tasks that need human attention can be outsourced en masse to the crowd with online access, giving much promise to researchers by significantly reducing the cost and time to obtain the micro-level annotations needed for their studies. When using crowdsourcing for micro-level multimedia annotations, two main challenges arise: evaluation and interface. While many evaluation metrics have already been proposed to assess the quality and agreement of macro-level annotations, there has been limited work to evaluate the agreement in the micro-level. There is also a need of web interfaces that allow crowd workers to accurately and efficiently annotate micro-level events and behaviors online.

2. HYPOTHESES & CONTRIBUTIONS

Broadly, the thesis will comprise of studies that investigate 3 main hypotheses:

H1: It is possible to use crowdsourcing to effectively obtain micro-level annotations of human behaviors in multimedia

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

ICMP'13, December 9–13, 2013, Sydney, Australia.

ACM 978-1-4503-2129-7/13/12.

<http://dx.doi.org/10.1145/2522848.2532198>

¹ <http://youtube-global.blogspot.fr/2013/05/heres-to-eight-great-years.html>

contents that are of quality comparable to those coded by local expert annotators.

H2: The communication modality affects the persuasiveness of speakers in social multimedia contents.

H3: In creating computational models of speakers' behaviors, multi-modality is important for machine classification of whether social multimedia contents are persuasive or not.

While pursuing to answer these hypotheses, the thesis will make the following specific contributions to the research community:

- Develop web-based annotation tools that allow precise and convenient multimedia behavior (or event) annotations, directly portable to popular crowdsourcing platforms such as Amazon Mechanical Turk.
- Confirm that crowdsourcing can be an effective way to obtain quality annotations of human behaviors.
- Introduce new evaluation procedures (based on Krippendorff's alpha) in measuring annotation agreement of human behaviors.
- Perform an in-depth perception study to investigate how communication modality affects the persuasiveness of speakers in social multimedia contents.
- Analyze and identify important behaviors in social multimedia contents that influence human perception of persuasiveness.
- Experiment with machine learning algorithms and computational descriptors of human behaviors for the machine to identify between persuasive and non-persuasive contents.

3. BACKGROUND

Crowdsourcing. Crowdsourcing has gained much attention lately, and a survey paper by Yuen et al. [25] and another by Quinn and Bederson [18] present a general overview of the topics on crowdsourcing and human computation. There are a number of studies on crowdsourcing video-related tasks [2, 20, 24], but many of them are on labeling videos at the whole level and are not concerned with micro-level annotations. Probably most relevant pieces of work in terms of web interfaces needed for crowdsourcing micro-level human behaviors were done by Vondrick et al. [23] and Spiro et al. [22], whose interfaces allow micro-level video annotations and are also used with Amazon Mechanical Turk. However, their interfaces put an emphasis on motion tracking, while this thesis requires an interface mainly concerned with annotating behavioral events in videos. Although there are quite a number of off-line software for making annotations on videos [7], such full-fledged tools are not suitable to be used for crowdsourcing due to a relatively steep learning curve and the difficulty in incorporating them into online crowdsourcing platforms like Amazon Mechanical Turk.

Annotation Agreement Metrics. Krippendorff's alpha has been previously used to measure inter-coder reliability of video annotations both at the macro-level [20] (label on the whole video level) and micro-level [10]. However, there is a need to further explore the stability of the alpha at different temporal resolutions, and more importantly, there are no metrics specifically designed

to fairly measure the agreement between coders for subtle micro-level and event-level behavior annotations.

Persuasiveness. Previous studies on nonverbal communication have identified the effect of several nonverbal cues on persuasiveness, but they were usually mediated by other factors and were context-dependent. The effect of speech rate on persuasion was studied in [3], and [11] found a positive correlation between gaze and compliance. Proximity is also shown to be influential on compliance [9]. Researchers in the field of human-computer interaction (HCI) have explored the effect of nonverbal cues on designing persuasive interfaces and found that adopting certain principles improves the persuasiveness of the machine [1, 6, 13]. Physical appearance is another factor which has been found to be influential in persuasion with attractive people perceived as more persuasive [4].

Communication modality is another factor shown to be effective in persuasiveness. In [5], it was shown that a likable speaker is more persuasive in video and audio versus text while an unlikable speaker is more persuasive in text.

4. RESEARCH PLAN

The research plan is comprised of three main parts to address each of the hypotheses outlined in Section 2. The first part is on crowdsourcing micro-level behavior annotations in multimedia contents, the second part on performing a perception study on persuasiveness in social multimedia contents, and the last part on analyzing human behaviors and experimenting with machine learning algorithms to investigate machine performance on identifying persuasive contents.

4.1 Crowdsourcing Quality Annotations

In order to use crowdsourcing for obtaining micro-level annotations of human behaviors in multimedia contents, the first requirement is an effective interface that allows crowd workers to make annotations online. For this purpose, web-based annotation tools will be developed to allow crowd workers to conveniently make precise micro-level human behavior annotations online, which will also be easily integrated and used with popular crowdsourcing platforms such as Amazon Mechanical Turk.

To confirm that it is possible to effectively obtain micro-level annotations of human behaviors that are of quality comparable to those coded by local expert annotators, experiments will be performed on Amazon Mechanical Turk to compare the annotation results by crowd workers and those by local expert coders. Furthermore, various procedures will be tested to improve effectiveness in obtaining crowdsourced annotations, such as using frame-based majority voting with repeated annotations and introducing a robust training phase.

Lastly, to address the problem of not having sufficient metrics to measure the agreement in micro-level behavior annotations, various evaluation metrics will be proposed and tested, specifically time-slice (or frame-based) Krippendorff's alpha and event-based Krippendorff's alpha.

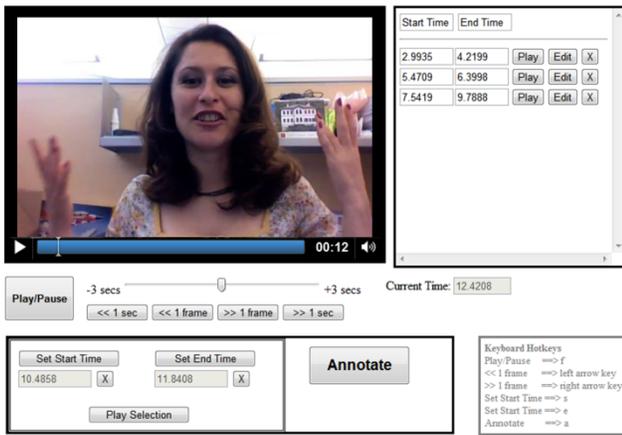


Figure 1. OCTAB (Online Crowdsourcing Tool for Annotations of Behaviors).

4.2 Perception Study on Persuasiveness

In order to fully understand human perception of persuasiveness in social multimedia contents, the first step would be to investigate how different modalities of communication influence our perception. Using movie review videos posted on YouTube, in which people give their opinions on movies, I plan to study the separate influence of three types of communication modalities of text, audio, and video. By understanding the relative importance of each modality in human perception of persuasiveness in social multimedia contents, one will be better able to design computational descriptors of human behaviors and machine learning algorithms for the machine to better identify between persuasive contents and non-persuasive ones.

4.3 Analysis of Persuasiveness in Social Multimedia Contents

Using the dataset of YouTube movie review videos gathered in the perception study, various verbal and nonverbal behaviors will be annotated using crowdsourcing, which will be analyzed to identify key factors that influence the level of persuasiveness in social multimedia contents. Then, the analysis results will be verified by creating computational descriptors of key behaviors and performing experiments with machine learning algorithms to test the machine performance of disambiguating between persuasive and non-persuasive contents.

In creating computational descriptors of behaviors, automatic annotations (and automatic feature extractions and descriptors) will also be studied separately with the vision of creating an automatic system that can parallel human performance in determining persuasiveness.

5. PRELIMINARY RESULTS

This section briefly explains the preliminary results to date, especially on the web-based annotation interface for crowdsourcing, the quality of crowdsourced annotations of micro-level human behaviors, and the perception study of persuasiveness in social multimedia contents.

5.1 Crowdsourcing Quality Annotations

I developed OCTAB (Online Crowdsourcing Tool for Annotations of Behaviors) [16], which is a web-based interface

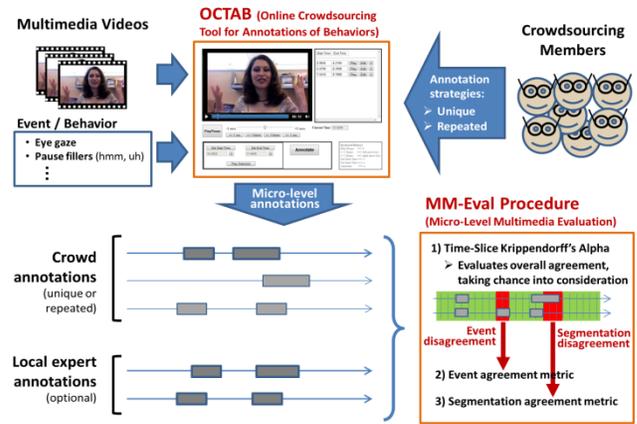


Figure 2. Overview of our approach for crowdsourcing micro-level video annotations, with a focus on our new interface, called OCTAB, and evaluation procedure for multimodal behavior annotations, called MM-Eval.

that allows crowd workers to conveniently navigate in a video to annotate micro-level human behaviors or events (see Figure 1). OCTAB is a light-weight, HTML-based interface that enables frame-level precision in making annotations. Its design offers high usability to improve convenience and accuracy, and it can also be easily ported and used with popular crowdsourcing platforms like Amazon Mechanical Turk.

As depicted in Figure 2, I also proposed a new evaluation procedure, called MM-Eval (Micro-level Multimedia Evaluation), which compares fine time-aligned (frame-based) annotations using Krippendorff's alpha metric and introduced two new metrics to evaluate the types of disagreement between coders. With an experiment using OCTAB and the evaluation procedure, preliminary results showed that a majority vote among repeated annotations from 3 crowd workers leads to a quality comparable to that of local expert annotations.

Currently, I am working on devising an event-based Krippendorff's alpha metric which will supplement time-slice (or frame-based) Krippendorff's alpha and on performing experiments to study how a robust training phase improves crowdsourced annotations.

5.2 Perception Study on Persuasiveness

Prior to the perception study on persuasiveness, various nonverbal behaviors have been studied in the context of negotiation [15, 17], which are expected to have relevance in the context of persuasiveness as well.

For the perception study on persuasiveness [14], a set of 86 movie review videos were collected from YouTube. After teasing apart each communication modality (text, audio, and video) separately, a preliminary experiment investigated the influence of each modality in perceived persuasiveness. The results revealed that modality does have different influence on the perception of persuasiveness, with visual modality having the most influence (Figure 3).

6. CONCLUSION

My research goal is in understanding human perception of persuasiveness in social multimedia contents. It involves studying how different communication modalities influence our perception

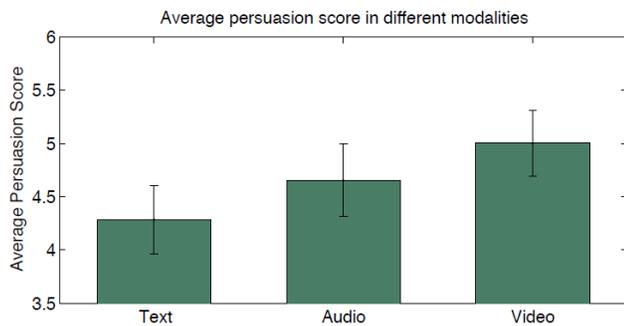


Figure 3. The average persuasiveness score in three communication modalities. Persuasiveness is significantly higher in video vs. text ($p < 0.001$).

and identifying key verbal and nonverbal behaviors that eventually lead us to believe someone is convincing and influential. In addition, researchers can effectively analyze human behaviors with crowdsourcing to obtain annotations of micro-level human behaviors, and this work will provide to the research community a convenient web-based tool, effective procedures for obtaining high-quality annotations from crowd workers, and evaluation metrics to fairly and accurately measure the quality and agreement of micro-level behavior annotations.

7. ACKNOWLEDGMENTS

Special thanks to Prof. Louis-Philippe Morency for his warm and passionate support in every stage of my Ph.D. endeavor.

8. REFERENCES

- [1] Bailenson, J., Beall, A., Loomis, J., Blasovich, J., and Turk, M. 2005. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Hum. Commun. Res.* 31, 4 (Oct. 2005), 511-537.
- [2] Biel, J. and Gatica-Perez, D. 2012. The good, the bad, and the angry: Analyzing crowdsourced impressions of vloggers. In *Proc. of the 6th Int'l AAAI Conf. on Weblogs and Social Media*. ICWSM '12, 407-410.
- [3] Buller, D., LePoire, B., Aune, R., and Eloy, S. 1992. Social perceptions as mediators of the effect of speech rate similarity on compliance. *Hum. Commun. Res.* 18, 2 (Dec. 1992), 286-311.
- [4] Chaiken, S. 1979. Communicator physical attractiveness and persuasion. *J. Pers. Soc. Psychol.* 37, 8 (Aug. 1979), 1387-1397.
- [5] Chaiken, S. and Eagly, A. 1983. Communication modality as a determinant of persuasion: The role of communicator salience. *J. Pers. Soc. Psychol.* 45, 2 (Aug. 1983), 241-256.
- [6] Chidambaram, V., Chiang, Y.-H., and Mutlu, B. 2012. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *Proc. of the 7th ACM/IEEE Int'l Conf. on Human-Robot Interaction*. HRI '12, 293-300.
- [7] Dasiopoulou, S., Giannakidou, E., Litos, G., Malasioti, P., and Kompatsiaris, Y. 2011. A survey of semantic image and video annotation tools. In *Knowledge-Driven Multimedia Information Extraction and Ontology Evolution*, G. Paliouras, C. Spyropoulos, and G. Tsatsaronis, Ed. Lecture Notes in Computer Science. Springer Berlin, 196-239.
- [8] Fogg, B. 2002. *Persuasive Technology: Using Computers to Change What We Think and Do* (Chapter 5, pg. 89-120). Morgan Kaufman Publishers, New York.
- [9] Glick, P., Demorest, J., and Hotze, C. 1988. Keeping your distance: Group membership, personal space, and requests for small favors. *J. Appl. Soc. Psychol.* 18, 4 (Mar. 1988), 315-330.
- [10] Kang, S., Gratch, J., Sidner, C., Artstein, R., Huang, L., Morency, L.-P. 2012. Towards building a virtual counselor: Modeling nonverbal behavior during intimate self-disclosure. In *Proc. of the 11th Int'l Conf. on Autonomous Agents and Multiagent Systems*. AAMAS '12, 63-70.
- [11] Kleinke, C. and Singer, D. 1979. Influence of gaze on compliance with demanding and conciliatory requests in a field setting. *Pers. Soc. Psychol. B.* 5, 3 (Jul. 1979), 386-390.
- [12] Miller, G. 1980. *Persuasion: New Directions in Theory and Research* (Chapter 1, pg. 11-28). Beverly Hills, CA: Sage.
- [13] Miller, G. 2007. The art of virtual persuasion. *Science* 317, 5843 (Sept. 2007), 1343.
- [14] Mohammadi, G., Park, S., Sagae, K., Vinciarelli, A., and Morency, L.-P. 2013. Who is persuasive? The role of perceived personality and communication modality in social multimedia. In *Proc. of the 15th ACM Int'l Conf. on Multimodal Interaction*. ICMI '13.
- [15] Park, S., Gratch, J., and Morency, L.-P. 2012. I already know your answer: Using nonverbal behaviors to predict immediate outcomes in a dyadic negotiation. In *Proc. of the 14th ACM Int'l Conf. on Multimodal Interaction*. ICMI '12, 19-22.
- [16] Park, S., Mohammadi, G., Artstein, R., and Morency, L.-P. 2012. Crowdsourcing micro-level multimedia annotations: The challenges of evaluation and interface. In *Proc. of the Int'l ACM Workshop on Crowdsourcing for Multimedia*. CrowdMM '12, 29-34.
- [17] Park, S., Scherer, S., Gratch, J., Carnevale, P., and Morency, L.-P. 2013. Mutual behaviors during dyadic negotiation: Automatic prediction of respondent reactions. In *Proc. of the 5th Biannual Humaine Assoc. Conf. on Affective Computing and Intelligent Interaction*. ACII '13.
- [18] Quinn, A. and Bederson, B. 2011. Human computation: A survey and taxonomy of a growing field. In *Proc. of the 2011 Annual Conf. on Human Factors in Computing Systems*. CHI '11, 1403-1412.
- [19] Reardon, K. 1991. *Persuasion in Practice*. Sage Publications, Inc., CA.
- [20] Riek, L., O'Connor, M., and Robinson, P. 2011. Guess what? A game for affective annotation of video using crowd sourcing. In *Proc. of the 4th Int'l Conf. on Affective Computing and Intelligent Interaction*. ACII '11, 277-285.
- [21] Segrin, C. 1993. The effect of nonverbal behavior on outcomes of compliance gaining attempts. *Communication Studies* 44, 3-4, 169-187.
- [22] Spiro, I., Taylor, G., Williams, G., and Bregler, C. 2010. Hands by hand: Crowd-sourced motion tracking for gesture annotation. In *Proc. of the 2010 IEEE Computer Vision and Pattern Recognition Workshops*. CVPRW '10, 17-24.
- [23] Vondrick, C., Ramanan, D., and Patterson, D. 2010. Efficiently scaling up video annotation with crowdsourced marketplaces. *Computer Vision - ECCV 2010*, 6314, 610-623.
- [24] Wu, S.-Y., Thawonmas, R., and Chen, K.-T. 2011. Video summarization via crowdsourcing. In *Proc. of the 2011 Annual Conf. Ext. Abst. on Human Factors in Comp. Sys.* CHI EA '11, 1531-1536.
- [25] Yuen, M.-C., King, I., and Leung, K.-S. 2011. A survey of crowdsourcing systems. In *Proc. of the IEEE 3rd Int'l Conf. on Social Computing*. SocialComp '11, 766-773.
- [26] Zimbardo, P. and Leippe, M. 1991. *The Psychology of Attitude Change and Social Influence*. McGraw-Hill, New York.