Are you friendly or just polite? – Analysis of smiles in spontaneous face-to-face interactions

Mohammed (Ehsan) Hoque[1], Louis-Philippe Morency[2], Rosalind W. Picard[1]

[1] MIT Media Lab, Cambridge, MA 02139, USA.
{mehoque, picard}@media.mit.edu
[2] Institute for Creative Technologies, University of Southern California,
Marina del Ray, CA 90292, USA.
morency@ict.usc.edu

Abstract. This work is part of a research effort to understand and characterize the morphological and dynamic features of polite and amused smiles. We analyzed a dataset consisting of young adults (n=61), interested in learning about banking services, who met with a professional banker face-to-face in a conference room while both participants' faces were unobtrusively recorded. We analyzed 258 instances of amused and polite smiles from this dataset, noting also if they were shared, which we defined as if the rise of one starts before the decay of another. Our analysis confirms previous findings showing longer durations of amused smiles while also suggesting new findings about symmetry of the smile dynamics. We found more symmetry in the velocities of the rise and decay of the amused smiles, and less symmetry in the polite smiles. We also found fastest decay velocity for polite but shared smiles.

Keywords: polite smiles, amused smiles, shared smiles, banking dataset, smile analysis.

1 Introduction

A smile is one of the simplest forms of expressions that is easy for humans to recognize. Several studies have reported success in developing computational models that can recognize smiles with fairly high accuracy [2]. Even though there has been a trend to equate smiles with the activation of lip corner pull (AU 12) and cheek raise (AU 6), several studies have attempted to disambiguate among different kinds of smiles (e.g., deliberate vs. genuine) by exploring their morphological and temporal patterns [1][5][8][9].

Being able to automate recognition of smiles has opened up new possibilities in areas such as conversational agents, customer service, and cognitive behavior modeling. However, a smile is a multi-faceted dynamic expression that can signal much more than "happy" – it can also indicate rapport, polite disagreement, sarcasm, frustration, pain and more. Even with one category of smile, there are ways to vary the dynamic and morphological properties of the smile to indicate the scale and sincerity of that expression. How are the properties of smiles different when they are

shared vs. solo? Previously, these kinds of questions have been very difficult to answer, especially since it is not trivial to collect large sets of labeled spontaneous expression data from quality-recorded natural conversational interactions.

In the past, Ambadar et al. [3] investigated morphological and dynamic properties of deliberate and genuine smiles. They collected data on a study where participants were brought to the lab to act various facial expressions. Between acting and data collection, the participants voluntarily looked at the experimenter and smiled. Those examples were then tagged by judges and were used to analyze the properties of deliberate and genuine smiles. In another study [4], Ochs et al. investigated the morphological and dynamic characteristics of amused, polite, and embarrassed smiles displayed by a virtual agent. A web platform was developed for users to provide smile descriptions (amused, polite and embarrassed) for a virtual agent. While these studies have been extremely useful to motivate the problem with initial exploratory results, none of them really address the issues of understanding those smiles in contextual face to face interactions when those smiles are shared and not shared.

In this study, we utilize a dataset collected by Kim et al. at MIT [10] [11], which contains spontaneous face-to face interactions in a banking environment, where smiles were labeled by both participants after the interaction. While the dataset is labeled for various expressions, for this study, we focus on understanding the differences between polite and amused smiles and how these change when smiles are shared or occur to just one participant. In particular, we focus more on understanding the difference in durations, occurrences, and dynamic properties of polite and amused smiles. The remaining part of the paper is organized as follows: Section 2 describes the dataset and experimental set up. Section 3 describes the research questions addressed in this paper, while Section 4 reports on the current findings. Section 5 provides discussions on the results and future work.

2 Spontaneous Face-to-Face Banking Dataset

This section describes how the data was collected and is largely an excerpt from the work of Kim [10]. In Kim et al.'s MIT study, young adults interested in learning about banking services were invited to meet with a professional banker in a conference room (Figure 1). The bankers provided information about two kinds of financial services just as they did at the retail branches where they worked during the day. The first service was to cash a $5 voucher from the participant as compensation for participating in the study. This part was designed to simulate a cashing a check scenario. The participants were recruited in the study with the incentive of getting $10 for their participation. However, after each arrived, the banker told him or her that they could only get $5 for now and would need to fill out additional paper work to claim their remaining $5. This manipulation was made to instill a slightly negative state in the customer in order to mitigate the "it's fun to be an experiment" phenomenon and also to approximate the real-world situation where a customer often goes to a bank feeling a little negative because of a need to fix a problem. After the experiment ended the participant received the rest of the money without additional paperwork.

The second service was for the banker to explain one of four financial services that a customer chose to learn more about. This part was designed to allow the customer to ask questions and receive information about the financial product just as they would in a real bank visit.

2.1 Participants

Two professional personal bankers were hired, each with over two years of career experience as a personal banker, to do what they usually do at work - explain financial services. One banker interacted with seventeen participants, while the other interacted with forty-four. Each experiment included one banker with one customer.

Before hiring, the bankers were asked if they would be willing and able to manipulate the type of facial expressions displayed during interaction with the customers. Each banker agreed to alter his facial expressions in three different ways, following these exact instructions.

- Manipulation 1 – Neutral facial expressions: Sustain neutral facial expressions over the entire interaction.
- Manipulation 2 – Always smiling: Sustain smiling over the entire interaction.
- Manipulation 3 – Complementary facial expressions, i.e., empathetic: Understand the customer's feeling and respond to it appropriately by smiling when the customer seems to feel good.

Throughout the experiment, the bankers interacted with the customer normally in addition to maintaining one of the three manipulations. This included greeting the customer, providing proper information, and thanking the customer for their time. The facial expressions and the voices of the banker and of the customer were unobtrusively recorded using a video camera from the moment they met and greeted to the end when the customer left. Customers were not told about the banker's facial expression manipulations and all the interactions appeared to proceed very naturally.

Forty one males and twenty females (n=61) who were interested in receiving information about different financial services were recruited through flyers. Before the experiment started, they were told that their face and voice data would be recorded as banks normally do for security reasons. However, they were not told that their facial expressions would be analyzed until after the study. Afterward, they were told about the need to analyze expressions and they were asked to help label them.

2.2 Experimental Setup

The experiment was conducted in a room equipped with a desk, two chairs, bank service advertising pamphlets and two cameras to make the appearance alike to a personal banking service section at banks (Figure 1). One camera was used to record the banker's facial expressions and the other was used to record the participant's facial expressions.

Prior to the participant entering the room the banker was told which expression manipulation to conduct. The participant was then allowed into the experiment room where they would interact with the banker and learn about specific financial services. At the end of the experimental interaction, which took about 10 minutes on average, both the banker and participant filled out 9-point Likert scale surveys evaluating the quality of the service based on the most comprehensive and popular instrument SERVQUAL [6] and the attitude of the banker.

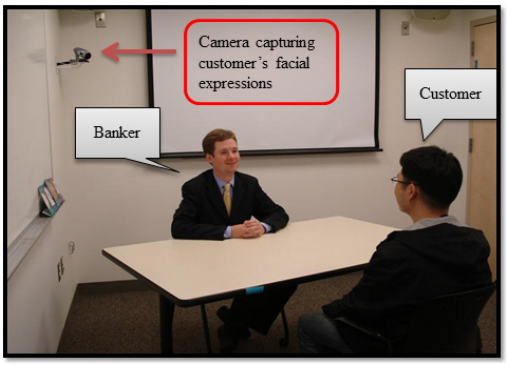

Fig. 1. Experimental set up for banker and the customer interaction. The camera that is visible behind the banker is capturing the facial expressions of the customer. There is another camera, not visible in the image, behind the customer, capturing the facial expressions of the banker.

2.3 Facial Expressions Coding

After the banker and participant finished the surveys, they were debriefed and asked to label the video data for their facial expressions. After labeling their own video information, they labeled the videos containing the person they interacted with (e.g., banker coded customer & customer coded banker). Therefore, for each conversation, there are two videos containing the facial expressions of banker and customer.

Bankers and customers used custom labeling software to label their expressions and affective states. The label interface contained two parts: the upper part displayed the video and the lower part provided the entity for the banker and the participant to enter the time when a certain facial expression was observed and seven affective labels to select. These seven labels were: smile, concerned, caring, confused, upset, sorry, and neutral. If there was no proper label to choose from, the user could press "Other" and enter another label that they thought was appropriate for the expression. The labelers were instructed to stop playing the video and click on the label button when they saw a facial expression, and then to continue to play the video until they saw a change in the facial expression. On the right side of the user interface, there was a text box displaying the time and the labeling result and it was editable so that the user could annotate the reason for each facial expression, e.g. “smile – he made me

laugh". The labelers were instructed to group every smile as either polite or amused. These extra labels were entered manually in the text box.

3 Research questions

In this paper, we focus primarily on understanding the differences between polite and amused smiles. We anticipate that polite smiles in the context of our dataset are more likely to be social, masking and controlled smiles while the amused smiles in the context of our dataset are more likely to be genuine, and felt. In this study, we focus our attention towards exploring the differences between polite and amused smiles in face to face interactions.

We are primarily interested to explore three questions in this study. When people exhibit amused and polite smiles in a task-driven spontaneous face-to-face interactions:

1) Are there any differences in terms of durations between polite and amused smiles?
2) Do amused and polite smiles get shared by the conversation partner? Do people share them equally or does one type get shared more often?
3) Are there any differences in dynamic features between polite and amused smiles? How can we quantify the difference of dynamic?

We are interested to motivate and gain further insights on these questions because understanding those aspects of human communication can help develop models (e.g., virtual human) that can naturally mimic face-to-face interactions with other humans.

4 Experiments

To directly address our research questions, we performed a series of experiments. This section describes our experimental setup and results from our deep analysis.

4.1 Smile Annotation and Segmentation

In our study, we do not measure the dynamics of the bankers' smiles since they were manipulated; we only analyze the dynamics of the customers' smiles and whether or not their smiles occurred in conjunction with a banker smile or solo.

As mentioned in section 2, each customer video was labeled for polite and amused smiles by the banker and by the customer him/herself. We did not use the labels produced by the customers since, after looking at them and seeing huge variation in how the labels seemed to be applied, we realized different customers seemed to interpret the labels differently. Using the two banker's labels led to significantly more consistent labels as judged by outward appearance of the expressions. We therefore chose to use the labels produced by the bankers which are more likely to be consistent. Using a third party coder to code the smiling instances is an option that we

are planning to implement in our future work. One significant advantage of using the banker's labels is that they are automatically taking conversational context into account when interpreting the smiles.

The labelers (bankers) indicated individual points in the video where they observed polite and amused smiles. Therefore, extra work was needed to be able to approximate the exact beginning and end points of each marked smile. Given the variability in the data, we manually annotated the beginning and end of each smile given the initial labels produced by the bankers. Through this process, we gathered 227 clips of amused smiles and 28 samples of polite smiles encompassing 61 participants playing the role of customers. We were also interested to find out which of those samples of smiling instances were also shared by the banker. Therefore, we separated the smiling instances of customers where the banker also self-labeled himself to be exhibiting the same kind of smiles (polite or amused).

4.2 Duration and Timing

The average duration of customers' shared polite, shared amused and unshared polite and unshared amused smiles are shown in Table 1.

Table 1. Comparison of durations for customers' polite and amused smiles

	Average duration	standard deviation
shared amused smiles (n=44)	6.1 sec.	4.6
non-shared amused smiles (n=183)	4.7 sec.	3.0
non-shared polite smiles (n=21)	3.7 sec.	1.2
shared polite smiles (n=10)	3.2 sec.	0.77

It is evident that the amused smiles are usually longer when shared as opposed to unshared. Comparatively, the durations of polite smiles are usually the same regardless of whether the smile is shared or not. The high standard deviation for un/shared amused smiles also indicates that the distribution of durations for amused smile is pretty widespread, as shown in Table 1.

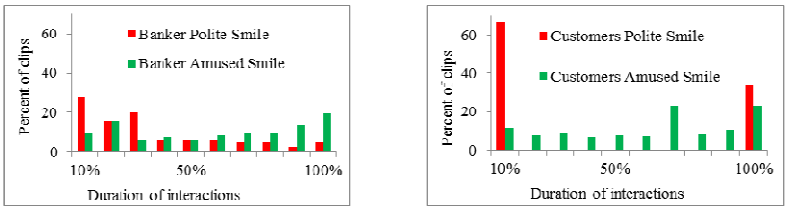

Fig. 2. Position of polite and amused smiles relative to the entire conversation. (a) Bankers yielded polite and amused smiles consistently throughout the interaction. (b) Customers yielded polite smiles only at the beginning and end of conversations, and amused smiles throughout the interaction.

We have also investigated the positions in respect to the entire conversations for amused and polite smiles, for both bankers and customers, as shown in Fig. 2. It is evident from Figure 2 that bankers seem to display polite and amused smiles throughout the interaction, whereas the customers seem to display polite smiles at the start and end of the conversations. On the other hand, about 1/3 of the 31 polite smiles were shared, while only about 1/5 of the 227 amused smiles were shared in these data.

4.3 Smile Dynamics: Rise, Sustain and Decay

Along with duration and position parameters, we were also interested to explore the dynamics of smile. We used the Sophisticated Highspeed Object Recognition Engine (SHORE) [7] API by Fraunhofer to detect the intensity/probability of smiles. The SHORE API provides a score between 0-100 for smiles by analyzing mouth widening, Zygomaticus muscles and other regions of the face in every frame, which creates a smile track per clip, as shown in Figure 3. We define three parameters to better analyze smile dynamics: rise, sustain and decay. Note that in our natural data, there was often not one clear "apex" or "peak" to the smile. Thus, we do not use the usual definition of onset time = "time to the highest peak", while, offset= "decay from that highest peak", because for spontaneous smiles, they often had a sustained region with multiple peaks, as in Fig. 3. Therefore, in this study, we refer to onset as rise time, offset as decay, and apex as sustain.

Careful observations indicated that the time stamps produced by the bankers were mostly the beginning of peak (L) of the smile without any further information on the rise and decay time as well as its sustain period.

The manual labeling process thus provided us with the beginning of rise times (R) and end of decay times (D). A visual example of where points L, R and R are more likely to be located is shown in Figure 3.

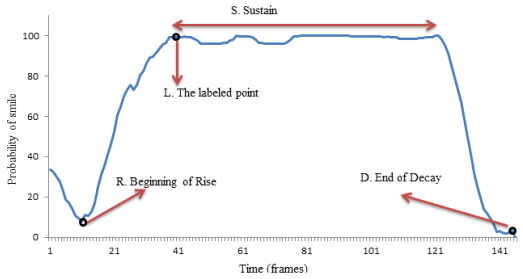

Fig. 3. A visual example of where points such as R (beginning of rise), D (end of decay) and S (sustain) could be located given the time stamp label, L, given by the labeler.

Our task was to automatically identify the region, S, which defined the time frame when the participants are more likely to be holding their smiles. We automated an algorithm to identify the locations where the probability of smiling is the highest. Then it traverses left and right looking for deviations that are higher than a pre-determined threshold to mark the start or end of the sustain period. For clips with

multiple peaks spread over the signal, the algorithm is biased towards selecting an initial point that is closer to the point labeled by the labeler.

Figure 4 provides the comparison of sustain period among shared polite/amused smiles and unshared polite/amused smiles. In these data we see that amused smiles have a longer sustain period than polite smiles. Additionally, shared amused smiles have longer duration for sustain compared to unshared amused smiles, whereas the duration of sustain for both shared polite and unshared polite is almost the same. This finding appears to be consistent with the popular notion that shared joy multiplies joy, here manifest by the extended duration of an amused smile.

In addition to sustain, we also analyzed the rise and decay times of amused and polite smiles, as shown in Figure 5. It is evident for both amused and polite smiles, regardless of whether they are shared or not, the difference between rise time and decay time is not statistically significant, and they are somewhat symmetric. Given this result, we decided to look more closely at the velocity for both rise and decay.

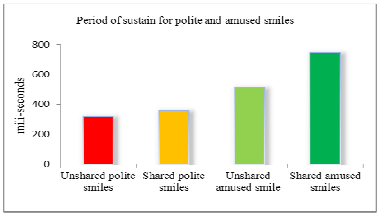

Fig. 4. Comparison of the period called sustain for (un)shared polite/amused smiles. The period of sustain for instances of shared amused smiles is the highest.

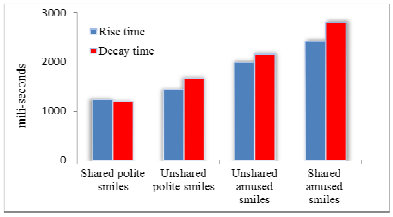

Fig 5. Comparison of rise, and decay time for (un)shared polite/amused smile instances. The ratio between rise time and decay time for all the categories seem very symmetrical.

4.4 Velocity of rise, sustain and decay

We analyzed the velocity of rise and decay signals for polite and amused smiles when they are shared vs. not shared. The velocity of rise (Vr) and decay (Vd) were defined as displacement in y axis divided by the elapsed time.

$$\text{Vr} = \frac{Y_s - Y_r}{T_s - T_r} \text{ and } \text{Vd} = \frac{Y_d - Y_s}{T_d - T_s}$$

where Y_s, Y_r and Y_d represent the smile intensity at the middle of the sustain period, the beginning of rise and at the end of decay, respectively. T_s, T_r and T_d represent the time at the middle of sustain, at the beginning of rise and at the end of the decay, respectively.

As shown in Figure 6, our analysis suggests that the amused smiles have the most symmetric velocities of rise and decay, whether shared or unshared, $V_d \approx V_r$. However, for polite smiles, these velocities were more asymmetric. Shared polite smiles decayed the fastest: $V_r < V_d$ while the polite smiles that rose the fastest were unshared $V_r > V_d$. As shown in figure 5, for shared and unshared polite smile

instances, the ratio between T_s - T_r (time difference between sustain and rise) and T_d - T_s (time difference between sustain and decay) remains almost same. It is the smile intensity (Y) that is contributing to the difference in velocities between shared polite and unshared polite instances.

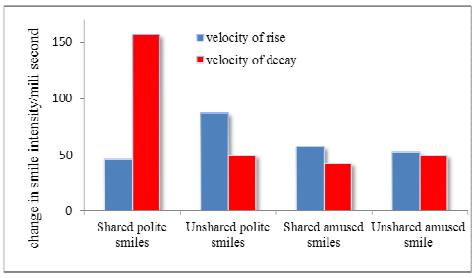

Fig. 6. Comparison of shared polite/amused smiles with unshared polite/amused smiles in terms of velocities

5 Discussion and Conclusions

In this study, we have investigated the phenomenon of polite and amused smiles in a new corpus of spontaneous face-to-face interactions. There were three key findings that have been reported in this paper. Our results suggested that duration of amused smiles are higher than the duration of polite smiles, which is consistent with what has been reported in the literature so far, although under different data gathering conditions. We additionally report that the duration of amused smiles are more likely to be higher when they are shared as opposed to solo. However, for polite smiles, the duration does not seem to change much regardless of whether the smile is shared or not (in fact, slightly higher duration when not shared).

In this spontaneous face-to-face banking dataset, we notice that when bankers labeled their polite and amused smiles during all the interactions, they seem to indicate that they have displayed polite and amused smiles consistently during the entire interaction, as shown in Fig. 2 (a). However, when the same banker labeled the corresponding customer video, he indicated the occurrences of polite smiles only in the beginning and end of the interactions, as shown in Fig. 2 (b). In other words, customers were viewed as less likely to share polite smiles with the bankers unless it happened at the beginning or end of the interactions. For amused smiles, the customers were more likely to share the smiles with the banker during the entire interaction. These data support a view that it is socially acceptable not to share a polite smile when it occurs in the middle of the discussion. Perhaps, we can argue that, in the context of a conversational agent, it may not be necessary or wise for the agent to mirror its user's smiles in every instance, but rather look for ones that are amused and "share" those.

One of the key findings in this paper is that amused smiles, whether shared or not, are more likely to be symmetrical in rise and decay velocities. We additionally report

that the duration of amused smiles are more likely to be higher when they are shared as opposed to solo. However, for polite smiles, the duration does not seem to change much regardless of whether the smile is shared or not.

In this paper, we report subtle differences of dynamic and morphological features between polite and amused smiles in face to face interactions. Hopefully, the reported findings will further motivate the development of automated systems that can differentiate between polite and amused smiles under natural conditions.

Acknowledgments. The authors gratefully acknowledge Kyunghee Kim who designed the experiment and collected the data used in this paper as part of her MS thesis. The authors would also like to acknowledge the participants for their time helping with this study and agreeing to have their data shared with other researchers. This research was partially supported by the National Science Foundation (NSF) IIS Award HCC-0705647 and by MIT Media Lab consortium sponsors.

References

1. J. F. Cohn, and K. L. Schmidt, "The timing of facial motion in posed and spontaneous smiles", *International Journal of Wavelets, Multiresolution and Information Processing, 2*, pp. 1-12, 2004.
2. J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, "Towards Practical Smile Detection", *Transactions on Pattern Analysis and Machine Intelligence*, pp. 2106-2111, November, 2009.
3. Z. Ambadar, J. F. Cohn, and L. I. Reed, "All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous", *Journal of Nonverbal Behavior*, 33 (1), pp. 17-34*,* 2009.
4. M. Ochs, R. Niewiadmoski, and C. Pelachaud, "How a virtual agent should "smile? Morphological and Dynamic Characteristics of virtual agent's smiles"**,** *Intelligent Virtual Agent Conference (IVA)*, Philadelphia, USA, 2010.
5. M. E. Hoque, R. W. Picard, "Acted vs. natural frustration and delight: Many people smile in natural frustration", 9th IEEE International Conference on Automatic Face and Gesture Recognition (FG'11), Santa Barbara, CA, USA, March, 2011.
6. A. Parasuraman, V. A. Zeithaml and L. L. Berry, "SERVQUAL: A multiple-iten scale for measuring consumer perceptions of service quality", Journal of Retailing, Vol. 4, pp. 12-39, 1988.
7. C. Kueblbeck and A. Ernst, "Face detection and tracking in video sequences using the modified census transformation", *Journal on Image and Vision Computing*, Vol. 24, Issue 6, pp. 564-572, 2006, ISSN 0262-8856.
8. E. Krumhuber, A. S. R. Manstead, and A. Kappas, "Temporal aspects of facial displays in person and expression perception: The effects of smile dynamics, head-tilt, and gender". *Journal of Nonverbal Behavior*, Vol. 31, pp. 39-56, 2007.
9. M.F. Valstar, H. Gunes and M. Pantic, "How to Distinguish Posed from Spontaneous Smiles using Geometric Features", in *Proceedings of ACM International Conference on Multimodal Interfaces (ICMI'07)*, pp. 38-45, Nagoya, Japan, November 2007.
10. K.Kim,"Affect Reflection Technology in Face-to-Face Service Encounters," MIT MS Thesis, September 2009.
11. K. Kim, M. Eckhardt, N. Bugg, R. W. Picard, "The Benefits of Synchronized Genuine Smiles in Face-to-Face Service Encounters", *IEEE International Conference on Computational Science and Engineering*, pp. 801-808, Vancouver, Canada, 2009.