2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII)

Ballistic Timing of Smiles is Robust to Context, Gender, Ethnicity, and National Differences

Maneesh Bilalpur
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, USA
mab623@pitt.edu

Saurabh Hinduja
Department of Psychology
University of Pittsburgh
Pittsburgh, USA
sah273@pitt.edu

Kenneth Goodrich
Department of Psychology
University of Pittsburgh
Pittsburgh, USA
kng22@pitt.edu

Jeffrey F. Cohn
Department of Psychology &
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, USA
jeffcohn@pitt.edu

***Abstract*—Smiles are highly variable. In some, contraction of the orbicularis oculi raises the cheeks and amplifies their intensity. In others, smile controls counteract the oblique pull of the zygomatic major, alter their shape, and decrease their intensity. Despite this variability, some features appear to be stereotypic. These features include a high correlation between the amplitude and velocity of smile onsets and same for smile offsets. The larger a smile's amplitude, the greater its velocity. This dependence is referred to as ballistic timing. In two relatively large publicly available databases (EB+ and Belfast), we tested the hypothesis of ballistic timing of smile onsets and offsets. We found high and consistent non-linear correlations between amplitude and velocity of both onsets and offsets that were robust to individual differences in persons (gender and ethnicity), context, presence or absence of the Duchenne marker, and country of residence (United States, Ireland, Peru). All R^2 exceeded 0.85. The findings were highly consistent with ballistic timing. They have implications for detecting smiles that may be posed (which have been found to violate ballistic timing) and for realistic synthesis of smiles in social robots and virtual humans. Smiles that depart from ballistic timing are likely to be perceived as false or uncanny.**

***Index Terms*—Social signaling, ballistic timing, smiles, virtual humans and social robots**

I. Introduction

Facial expressions play a key role in human communication. They communicate emotion, intention, individual differences, action readiness, group affiliation, and a range of other meanings. Among facial expressions, smiles (zygomatic major contraction, AU 12 in FACS [1], [2]) are among the most frequent. Differences in appearance and timing among smiles communicate enjoyment [3], embarrassment [4], [5], triumph [6], dominance or affiliation [7], intense pain [8], and negative affect [9]. Despite such variability, are there consistent signals in the smile display that are robust to context, emotion, intention, and individual differences in people?

From an evolutionary perspective, evolved social signals result from selective pressure for conspicuous, stereotyped, and redundant communication [10]. Stereotyped signals afford reliability of communication and are easily recognized across varied contexts and persons. Redundant and conspicuous signals maximize the likelihood that signals will be recognized and ensure their efficiency in communication [10]. Are there features of smiles that are consistent with evolved social signals [11]?

Smile onset, defined as the largest continuous increase in the displacement of the lip corners, is a prime candidate. [12] found that observers perceive maximum differences in happiness intensity within a third of a second, which is far less time than required to perceive the ultimate intensity of the display. The velocity of the movement was likely sufficient to signal the smile's ultimate intensity. This finding among others suggests that smile amplitude and velocity are tightly coupled in automatic movements. From amplitude or velocity one may infer the other.

The timing of automatic movement is referred to as ballistic timing. Ballistic timing is a widely-occurring biological constraint that has been observed across vertebrates, including ants [13], crickets [14], salamanders [15], and humans [16], [17]. We tested the hypothesis that ballistic timing holds for smile onsets and offsets and is robust to gender and ethnicity, context, and national differences. Specifically, we address the following questions:

1) Are *onset* amplitude and velocity of smiles highly correlated?
2) Are *offset* amplitude and velocity of smiles highly correlated?
3) Are these correlations robust to individual differences in gender, ethnicity, presence of the Duchenne (i.e., cheek raise, or AU 6 in FACS), context, and country. Countries included the U.S., Ireland, and Peru.

II. Related work

Smiles consist of an onset, one or more peaks, and an offset. Onsets and offsets have been detected by both manual inspection (e.g., using FACS [2]) and automatic measurement. Automatic measurement has included electromyography [18] and computer-vision based approaches [17], [19]. Smile peaks have been detected automatically as well [20]. We focus on onsets and offsets of spontaneously occurring (i.e, not posed) smiles.

Research reported in this publication was supported in part by the US National Institutes of Health under Award Number MH096951. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

In ballistic timing, velocity and amplitude are highly correlated. In an early study, Schmidt and Cohn [17] found strong evidence for ballistic timing of smile onsets and offsets in both social and solitary contexts. While amplitude and velocity differed markedly between contexts (smiles in a social context were more intense), the non-linear correlation between velocity and amplitude was similarly high in both contexts. Similarly, while offset duration was longer in Duchenne smiles than in non-Duchenne smiles, the non-linear correlation between velocity and amplitude was again comparably strong. Thus, while contexts differed in amplitude and duration, the strong non-linear correlation between amplitude and velocity remained stable.

Cohn et al. [18] studied the relation between amplitude and duration during the onset phase of spontaneous and posed (also referred to as deliberate) smiles. Spontaneous smiles are automatic movements, while posed ones are not. Spontaneous but not posed smiles evidenced a strong relation between onset amplitude and duration, consistent with ballistic timing. This difference in timing was sufficient to discriminate between spontaneous and posed smiles with over 90% accuracy.

Differences in the morphology and dynamic characteristics of certain smiles, such as those perceived as amused, polite, and embarrassed have been explored [4]. Of these smile types, smiles perceived as amused more often had longer duration, larger amplitudes, higher maximum velocity, and more abrupt onsets and offsets. Smiles perceived as polite were smaller in duration than those perceived as embarrassed. Perceived polite smiles tended to have smaller amplitude, were more likely to have brief duration, and tended to have gradual onset and offset. Schmidt et al. [19] found consistent differences in velocity and intensity between Duchenne and non-Duchenne smiles.

More recently, Hoque et al. [21] have shown differences in the duration of smiles and how that is affected by social usage. They controlled the nature of smiles expressed by employees during their interactions with customers. Employees were instructed to express either neutral expressions or polite smiles throughout the interaction or empathetic amused smiles when they fit the situation. They found that amused smiles were longer than polite smiles. When shared and non-shared smiles were compared, shared amusement smiles were longer than non-shared amusement smiles.

These studies suggest that while smiles differ in amplitude, duration, and velocity in relation to their perceived meaning (e.g., embarrassed versus enjoyment) and context, their timing with respect to the correlation of velocity and amplitude in both onset and offset phases remains highly consistent. Nevertheless, evidence for ballistic timing across different contexts and persons remains tentative. The number of participants in these studies has been relatively small; individual differences in gender, ethnicity, and nationality have been unexplored; and robustness of ballistic timing to differences in emotional context have been neglected. To address the limitations, we turned to two relatively large and varied databases to test the hypothesis that ballistic timing in smile onsets and offsets is robust to individual differences in participants, context, and country.

III. Materials

A. Datasets

Generalization of statistical studies is crucial towards understanding robust (and possibly universal) mechanisms involved in human interactions. Open-source datasets play a big role in performing such studies. However, the open-source datasets differ in terms of their experimental design, data collection, subjects' diversity and availability of reliable manual annotations. Databases such as DISFA [22], CK+ [23], UNBC-McMaster [24] designed to induce emotions are limited in terms of the sample size or the range of contexts to study the effects across context and demographic factors. Despite such challenges, we use two different relatively large databases to analyze the timing ofsmiles. Both databases contain subjects of diverse ethnicity in tasks intended to elicit different emotions. The choice of database is motivated to capture the effects of gender, ethnicity and context on the temporal dynamics of smiles at scale.

TABLE I: Distribution of subjects in EB+ and Belfast.

Subjects	**N**
EB+	195
Belfast	153
EB+ percent female	58.80%
Belfast percent female	42.40%
EB+ Non-Caucasian	51.20%
Belfast Non-Caucasian	17.70%

TABLE II: Distribution of smiles by context in EB+ and Belfast.

	Onset	**Offset**
EB+		
Duchenne		
Anger	18	25
Embarrassment	86	85
Fear	64	60
Amusement	105	105
Non-Duchenne		
Anger	13	6
Embarrassment	13	14
Fear	10	14
Amusement	12	12
Belfast		
Fear	103	103
Amusement	140	140

The EB+ database [25] consists of 195 subjects of diverse ethnicity (Caucasian, Asian, African American, Hispanic, Middle-Eastern and Indian) and includes both male and female subjects. We focused on four contexts: 1) Talking to an experimenter while they tell a joke; 2) Being insulted by the experimenter; 3) Playing a silly game; and 4) Anticipating physical threat. These contexts were intended to elicit amusement, anger, embarrassment, and threat, respectively. The intended emotions may or may not have been induced.

TABLE III: Description of contexts in EB+ and Belfast.

Context	EB+	Belfast
Anger	Experience harsh insults from the experimenter.	N/A
Embarrassment	Play a game to improvise a silly song.	N/A
Fear	Anticipate and experience a physical threat.	Retrieve a sweet from a box with warning signs.
Amusement	Talk to the experimenter and listen to a joke.	Watch a comedy routine in the presence of the experimenter.

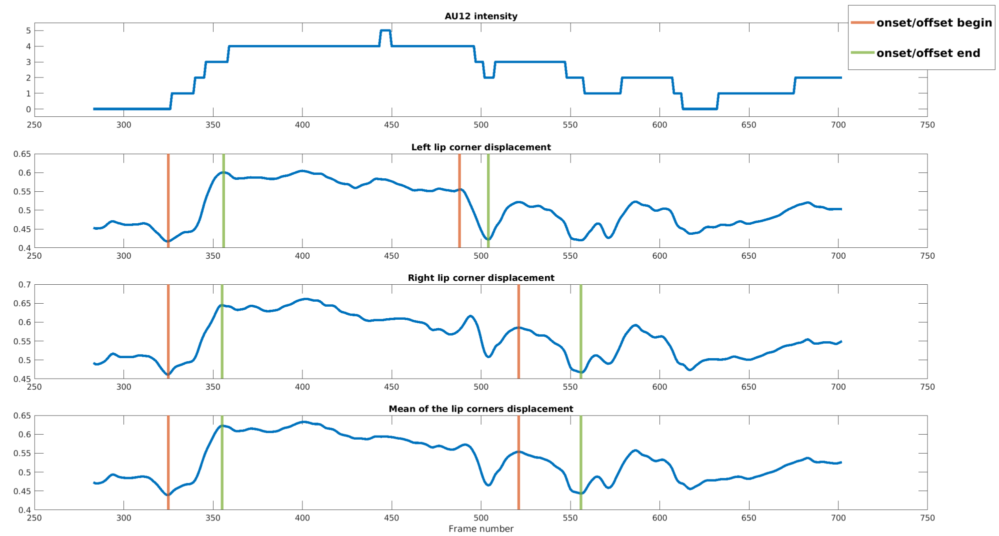


Fig. 1: Top to bottom: Figure showing AU 12 intensity annotation and automatically tracked left, right, and mean lip corner displacements. The initial frames of onset and offset phases are delineated with red lines. The final frames of the onset and offset phases are delineated with green lines.

For simplicity, we refer to these contexts by the emotions they were intended to elicit whether or not they did.

The Belfast Induced Natural Emotion Database [26] (here on referred to as Belfast) also contains contexts intended to elicit various emotions either through either active or passive tasks. An example of a passive task would be watching a video clip. As with EB+, we refer to contexts with respect to the emotions they were intended to elicit.

Each of the contexts (referred to as *tasks* in EB+ and *emotion targeted* in Belfast) was intended to elicit the target emotion, which may or many not have happened. The Belfast database was originally divided into three different sets depending on ethnicity of the subjects, sociality of the tasks and the targeted emotions. Because we are interested in ethnic differences of smiling, we combine the sets–1&3 (set-1 contains recordings of social amusement context collected in Ireland and set-3 contains recordings of social fear and amusement collected in both Ireland and Peru). This would provide us subjects from different ethnicities in social contexts. Table-III provides a brief summary of tasks involved in each context in EB+ and Belfast databases used in our investigation. In EB+, smiles were included if intensity was B level or higher in intensity and AU 4 (corrugator contraction) and AU 9 (nose wrinkling) were absent. 146 out of 195 subjects met this inclusion criteria. Because smile intensity and co-occurring actions (e.g., Duchenne marker) were not differentiated in the Belfast database, all smiles were included regardless of intensity and co-occurring actions.

We present the distribution of subjects and their demographics in EB+ and Belfast in Table-I. The distribution of smiles across contexts, and presence of the Duchenne marker across both databases are tabulated as Table-II.

B. Definition of a smile

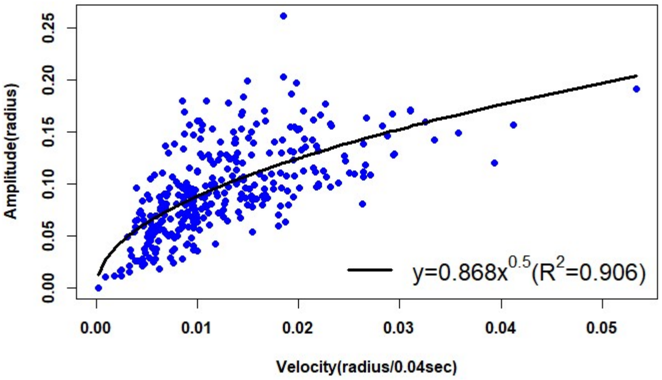
Onset is defined as the largest continuous increase in lip corner displacement. Offset is similarly defined as the largest continuous decrease in lip corner displacement. Unlike [17]–[19], onsets and offsets could correspond to the same or different peaks within the same smile. An example can be seen in figure-1.

We detect the onset, peak, and offset of each smile using an automated approach to lip corner tracking. This is similar to some of the existing works [17]–[19] and described in detail in section-III-D.

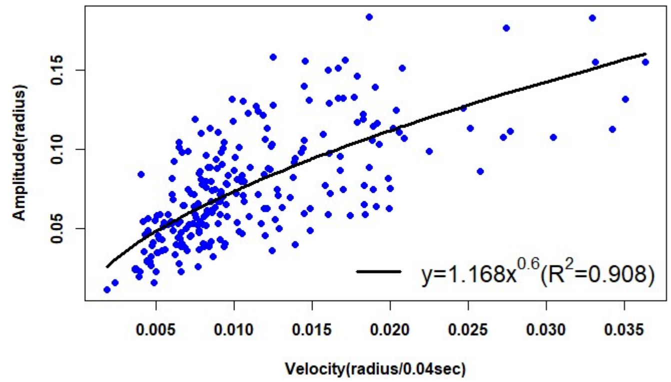
C. Duchenne smiles

Duchenne smiles [27] are comprised of AU 12 (Zygomatic major activation), which raises the lip corners obliquely and AU 6 (Orbicularis oculi activation), which raises the cheeks, narrows the eye aperture, and may cause wrinkling lateral to the eye corners.

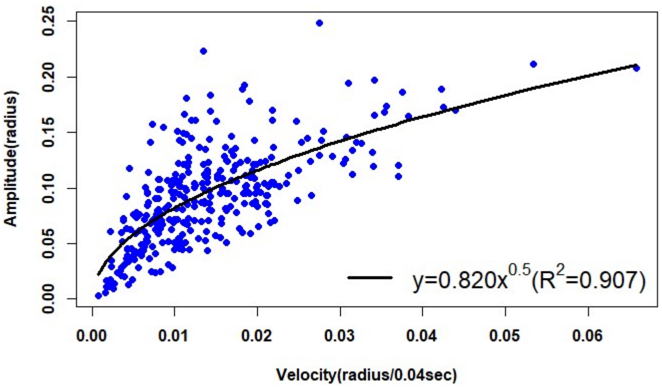
In FACS, AU intensity can vary on a five-point ordinal scale from A (trace level) to E (extreme). Smiles were defined by AU 12 intensity of B or higher. The Duchenne marker was defined similarly, as AU 6 intensity of B or higher. Because



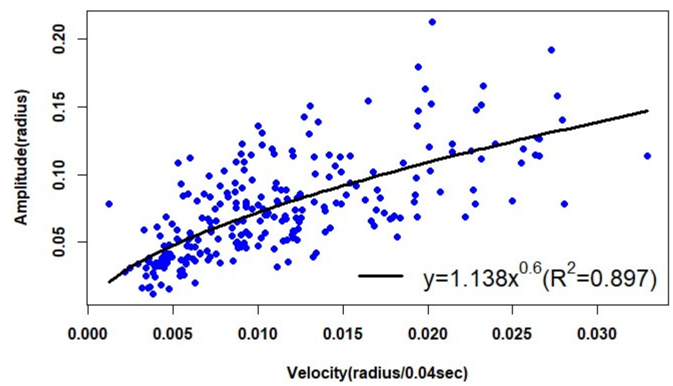
(a) Relation between amplitude and velocity in smile onsets in EB+. Slope is significant at $p < 0.001$.



(b) Relation between amplitude and velocity in smile onsets in Belfast. Slope is significant at $p < 0.001$.



(c) Relation between amplitude and velocity in smile offsets in EB+. Slope is significant at $p < 0.001$.



(d) Relation between amplitude and velocity in smile offsets in Belfast. Slope is significant at $p < 0.001$.

Fig. 2: Ballistic timing in smiles captured during onset and offset phases in EB+ (left) and Belfast (right). Each figure presents the curvilinear regression formulation and resultant explained variance (R^2).

the onset and offset phases for a given smile may be separated by more than one peak, the onset and offset phases of a smile may differ with respect to inclusion of the Duchenne marker.

D. Landmarks and displacement

To identify left and right lip corners, we use AFARtoolbox [28]–[30]. Lip corners are corrected for rigid motion (rotation, translation and scaling) and projected to 2D coordinates. We identify the mouth center as the mid-point of the left and right lip corners in the first frame of the AU annotated frames. We define the mouth width as the Euclidean distance between the left and the right lip corners from the same first frame. Displacement of a lip corner is measured as the Euclidean distance between the mouth center and the lip corner. Following the convention from [18] and [19] we normalize the lip corner displacements with the width of the mouth to cater to subjects of different shapes. All normalized lip corner displacements are smoothed using 4523H smoothing using R's *sleekts* package[1].

[18] used the right lip corner and [19] used the mean of the lip corners to capture the temporal progress of smiles. All these works have primarily characterised the temporal progress using three dynamics–*duration, amplitude* and *maximum velocity* (also referred as speed [18]). We use the mean of the right and left lip corner displacements to identify smiles (onsets and offsets) and extract their temporal dynamics.

[1]https://cran.r-project.org/web/packages/sleekts/sleekts.pdf

E. Temporal Dynamics

a) Duration: Time taken for the largest continuous lip corner displacement increase. Similarly for offsets, it is defined as the time taken for a smile for the largest continuous lip corner displacement decrease. We measure duration in seconds.

b) Amplitude: The largest continuous increase (or decrease) in the lip corner displacement during onset (or offset).

c) Maximum velocity: Maximum displacement between successive frames during onset (similarly for offset). This is also referred to as velocity (or speed in [18]).

Temporal dynamics involves three (duration, amplitude and maximum velocity) characteristic measures of smiles. In this study, we focus on amplitude and velocity because of their relevance to the ballistic timing.

TABLE IV: Explained variance(R^2) across contexts for the ballistic timing in EB+ and Belfast. All regression slopes are significant at p <0.001. Notice that Belfast does not have anger and embarrassment contexts and hence the N/A.

	Onset				Offset			
	Anger	Embarrassment	Fear	Amusement	Anger	Embarrassment	Fear	Amusement
EB+	0.934	0.922	0.899	0.896	0.929	0.924	0.902	0.905
Belfast	N/A	N/A	0.896	0.921	N/A	N/A	0.889	0.907

TABLE V: Explained variance(R^2) across gender for the ballistic timing in EB+ and Belfast. All regression slopes are significant at p <0.001.

	Onset		Offset	
	Male	Female	Male	Female
EB+	0.900	0.909	0.906	0.912
Belfast	0.903	0.914	0.894	0.903

TABLE VI: Explained variance(R^2) across ethnicities for the ballistic timing in EB+ and Belfast. All regression slopes are significant at p <0.001.

	Onset		Offset	
	Caucasian	non-Caucasian	Caucasian	non-Caucasian
EB+	0.917	0.904	0.910	0.903
Belfast	0.900	0.932	0.902	0.888

TABLE VII: Explained variance(R^2) across expressions for the ballistic timing in EB+ and Belfast. All regression slopes are significant at p <0.001. Notice that Belfast does not have AU6 annotations and hence the N/A.

	Onset		Offset	
	Duchenne	Non-Duchenne	Duchenne	Non-Duchenne
EB+	0.910	0.867	0.914	0.892
Belfast	N/A	N/A	N/A	N/A

IV. Results

Our goal is to evaluate ballistic timing in relation to context, gender, ethnicity, and presence of the Duchenne marker. We first perform a curvilinear regression to evaluate the hypothesis of ballistic timing. We then evaluate whether timing varies with context, gender, ethnicity or presence of the Duchenne marker. As noted above, because the Duchenne marker was not annotated in the Belfast database, comparisons of Duchenne and non-Duchenne smiles were limited to EB+ (refer section-III-C for the definition).

A. Evaluation of ballistic timing

We test for ballistic timing in both smile onset and offset phases. Ballistic timing refers to a strong positive relation between smile amplitude and velocity. Following [17], we use curvilinear regression to test the hypothesis of ballistic timing.

a) ***EB+:*** We present the relation between amplitude and velocity for onset phase and offset phase as figures-2a and 2c respectively. We notice that the curvilinear regression model in the onset phase ($y = 0.868x^{0.5}$) has a high explained variance ($R^2 = 0.906$) suggesting a strong association between amplitude and velocity. The explained variance ($R^2 = 0.907$) for the offset phase curve ($y = 0.820x^{0.5}$) is similar to that for the onset phase.

b) ***Belfast:*** In the onset phase, the amplitude-velocity relation is captured by a $y = 1.168x^{0.6}$ power curve with high explained variance $R^2 = 0.908$. This can also be observed in offset phase at $y = 1.138x^{0.6}$ with $R^2 = 0.897$. The corresponding regression plots are figures-2b and 2d.

We notice that the range of explained variance (R^2) is similar across both datasets. Similar to [17] we notice that both the onset and offset phases have similarly high explained variance in both EB+ and Belfast.

B. Effect of gender, ethnicity, context and Duchenne smiles on ballistic timing

Using the same curvilinear regression as above, we evaluate the effect of context, gender, ethnicity and expression. We perform regression for each level of these factors and use the correlation ($r = \sqrt{R^2}$) derived from the corresponding explained variance to compare shared variance between amplitude and velocity between levels. We perform Fisher‘s $r - to - Z$ transformation and significance test using the cocor [31] package for independent groups to evaluate the differences in correlation between the levels.

We hypothesize that ballistic timing is consistent across gender, ethnicity, context and the presence of Duchenne marker.

1) ***Gender:*** We evaluate the effect of gender on the ballistic mechanisms between amplitude and velocity through curvilinear regression models for each level (male or female). Table-V shows that the timing is prevalent across genders in both EB+ and Belfast. Fisher‘s transform does not reveal significant differences along gender in onset phase ($z = -0.365, p = 0.715$) and offset phase ($z = -0.288, p = 0.773$).

2) ***Ethnicity:*** Ballistic timing is consistent across differences in ethnicity (see table-VI)– both Caucasians and non-Caucasians have a significantly ($p < 0.001$) positive relation between amplitude and velocity. No significant differences were found between Caucasians and non-Caucasians for onset phase ($z = 0.693, p = 0.488$) and offset phase ($z = 0.375, p = 0.707$).

3) ***Context:*** The variance captured by context of smile origin for the ballistic timing is shown as table-IV. Anger smiles have the highest variance followed by embarrassment, fear, and amusement. Though we notice differences in the shared variance between contexts, we did not find any significant differences using the Fisher‘s transform. This trend is observed in both onset phase and offset phase. In Belfast, the shared variance is higher in amused smiles than fear smiles. This trend is observed in both onset phase and offset phase. However, the differences are not significant.

4) ***Duchenne:*** Presence of the Duchenne marker does not affect the shared variance (table-VII) captured by ballistic timing when compared to non-Duchenne smiles. Duchenne smiles have higher shared variance than non-Duchenne smiles. However, the differences are not significant in onset phase ($z = 1.296, p = 0.195$) and offset phase ($z = 0.753, p = 0.451$).

V. Discussion

Smiles are among the most common expressions, are highly variable in morphology and dynamics, and serve varied functions [3]–[5], [7], [9], [10]. As but one example, smiles that communicate amusement have longer duration, larger amplitude, and greater velocity than ones that are perceived as polite [4]. Despite a spectrum of such differences in smiles, our findings suggest that ballistic timing is common to all of them. Originally observed by Schmidt et al. [17], ballistic timing refers to the existence of a strong correlation between amplitude and maximum velocity of their onset and offset phases.

We evaluated the robustness of ballistic timing to a wide range of factors in two relatively large and diverse databases. Contexts included emotion inductions for anger, fear, embarrassment and amusement. Participants included men and women, Caucasians and non-Caucasians, and people from developed (US and Ireland) and developing (Peru) countries. Smiles with and without the Duchenne marker were included. We found strong evidence of ballistic timing that was robust to all potential sources of variation considered. In light of ubiquitous evidence for ballistic timing in non-human vertebrates [13]–[15], ballistic timing of smile onsets and offsets may represent a human universal. To test this hypothesis, further research would be needed.

The implications of our work can be observed in the fields of social robotics and intelligent virtual agents. Social robots and virtual agents can now be found helping children learn [32], assisting the elderly [33], promoting mental health discussion [34] and as multimodal conversational agents [35]. Mimicking human behavior by such agents is a key component to their success as it leads to increased human affinity [36]. However, insufficient emulation of human behavior by such agents could result in decreased human affinity towards them. The uncanny valley [36] would suggest that as we increase the degree of emulation, the affinity towards the agent decreases beyond certain level and then starts to increase. Though it is possible to design agents with reasonably safe level of affinity by pursuing a non-human design, we believe that in order to achieve the maximum possible affinity one needs to understand and incorporate fundamental mechanisms like the ballistic timing. The synthesis of gestures and facial expressions has become increasingly data-driven [37]–[39] because of neural networks. However, it is unclear if such data-driven approaches capture the invariant stereotypical mechanisms such as the ballistic timing in smiles. We believe that emulating such mechanisms would contribute to overcoming the uncanny valley and increase the humanness and user experience of social robots and virtual agents.

VI. Conclusions

Our study expands on earlier works where differences in smile expressions were studied largely in terms of limited contexts (solitary vs. social and spontaneous vs. posed) and limited variation in participants. We explore ballistic timing further in two large-scale publicly-available datasets. We consider the amplitude-velocity relation among smile onsets and offsets in relation to context, presence or absence of the Duchenne marker, gender, ethnicity, and country. We found strong evidence for ballistic timing robust to all these potential sources of variation. Our findings have implications for intelligent virtual agents and social robotics. In both domains, failure to implement ballistic timing could contribute to uncanny effects. By implementing ballistic timing of onsets and offsets of smiles and other displays as well, risk of uncanny valley effects might be reduced. Further, attention to ballistic timing could help identify deep-fakes to prevent harmful/unlawful impersonations of individuals.

Ethical Impact Statement

Our primary goal was to evaluate ballistic timing in smiles and its robustness to context, presence of the Duchenne marker, and individual differences in participants (gender, ethnicity, and country). Data were from two relatively large publicly-available data sources. All participants gave informed consent to use of their data by qualified investigators. Our findings if used in ethical manner could contribute to realistic animations and visual synthesis of smiles in virtual humans and social robots. A potential risk is that deepfakes could be made more realistic by making use of our findings.

References

[1] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

[2] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial Action Coding System: Facial action coding system: the manual: on CD-ROM*. Research Nexus, 2002.

[3] P. Ekman, "Basic emotions. handbook of cognition and emotion. 1999," in *Conference on Automotive User Interfaces and Interactive Vehicular Applications. https://ci. nii. ac. jp/naid/10028162696/ Portsmouth, NH, USA*, 1999, pp. 51–58.

[4] 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*, vol. 33, no. 1, pp. 17–34, 2009.

[5] D. Keltner and B. N. Buswell, "Embarrassment: its distinct form and appeasement functions." *Psychological bulletin*, vol. 122, no. 3, p. 250, 1997.

[6] L. F. Barrett, B. Mesquita, and M. Gendron, "Context in emotion perception," *Current Directions in Psychological Science*, vol. 20, no. 5, pp. 286–290, 2011.

[7] P. M. Niedenthal, M. Mermillod, M. Maringer, and U. Hess, "The simulation of smiles (sims) model: Embodied simulation and the meaning of facial expression," 2010.

[8] K. Craig, K. Prkachin, and R. Grunau, "The facial expression of pain. handbook of pain assessment," *New York: Guilford*, 2001.

[9] D. S. Messinger, W. I. Mattson, M. H. Mahoor, and J. F. Cohn, "The eyes have it: making positive expressions more positive and negative expressions more negative." *Emotion*, vol. 12, no. 3, p. 430, 2012.

[10] K. L. Schmidt and J. F. Cohn, "Human facial expressions as adaptations: Evolutionary questions in facial expression research," *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, vol. 116, no. S33, pp. 3–24, 2001.

[11] A. J. Fridlund, *Human facial expression: An evolutionary view*. Academic press, 2014.
[12] C. M. Leonard, K. K. Voeller, and J. M. Kuldau, "When's a smile a smile? or how to detect a message by digitizing the signal," *Psychological Science*, vol. 2, no. 3, pp. 166–172, 1991.
[13] S. Patek, J. Baio, B. Fisher, and A. Suarez, "Multifunctionality and mechanical origins: ballistic jaw propulsion in trap-jaw ants," *Proceedings of the National Academy of Sciences*, vol. 103, no. 34, pp. 12 787–12 792, 2006.
[14] R. Hustert and M. Baldus, "Ballistic movements of jumping legs implemented as variable components of cricket behaviour," *Journal of Experimental Biology*, vol. 213, no. 23, pp. 4055–4064, 2010.
[15] J. L. Van Leeuwen, "Why the chameleon has spiral–shaped muscle fibres in its tongue," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 352, no. 1353, pp. 573–589, 1997.
[16] S. Gibet, J.-F. Kamp, and F. Poirier, "Gesture analysis: Invariant laws in movement," in *International Gesture Workshop*. Springer, 2003, pp. 1–9.
[17] K. L. Schmidt, J. F. Cohn, and Y. Tian, "Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles," *Biological psychology*, vol. 65, no. 1, pp. 49–66, 2003.
[18] 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*, vol. 2, no. 02, pp. 121–132, 2004.
[19] K. L. Schmidt, Z. Ambadar, J. F. Cohn, and L. I. Reed, "Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling," *Journal of nonverbal behavior*, vol. 30, no. 1, pp. 37–52, 2006.
[20] F. D. l. Torre, T. Simon, Z. Ambadar, and J. F. Cohn, "Fast-facs: A computer-assisted system to increase speed and reliability of manual facs coding," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 57–66.
[21] M. Hoque, L.-P. Morency, and R. W. Picard, "Are you friendly or just polite?–analysis of smiles in spontaneous face-to-face interactions," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 135–144.
[22] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "Disfa: A spontaneous facial action intensity database," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.
[23] P. Lucey, J. F. 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 *2010 ieee computer society conference on computer vision and pattern recognition-workshops*. IEEE, 2010, pp. 94–101.
[24] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, "Painful data: The unbc-mcmaster shoulder pain expression archive database," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2011, pp. 57–64.
[25] I. O. Ertugrul, J. F. Cohn, L. A. Jeni, Z. Zhang, L. Yin, and Q. Ji, "Cross-domain au detection: Domains, learning approaches, and measures," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–8.
[26] I. Sneddon, M. McRorie, G. McKeown, and J. Hanratty, "The belfast induced natural emotion database," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 32–41, 2011.
[27] P. Ekman, R. J. Davidson, and W. V. Friesen, "The duchenne smile: Emotional expression and brain physiology: Ii." *Journal of personality and social psychology*, vol. 58, no. 2, p. 342, 1990.
[28] I. O. Ertugrul, L. A. Jeni, W. Ding, and J. F. Cohn, "Afar: A deep learning based tool for automated facial affect recognition," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–1.
[29] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3d face alignment from 2d videos in real-time," in *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, vol. 1. IEEE, 2015, pp. 1–8.
[30] ——, "Dense 3d face alignment from 2d video for real-time use," *Image and Vision Computing*, vol. 58, pp. 13–24, 2017.
[31] B. Diedenhofen and J. Musch, "cocor: A comprehensive solution for the statistical comparison of correlations," *PloS one*, vol. 10, no. 4, p. e0121945, 2015.
[32] Moxie, "Meet Moxie, a Social Robot That Helps Kids With Social-Emotional Learning," https://spectrum.ieee.org/moxie-a-social-robot-for-childhood-development, 2022, [Online; accessed 11-August-2022].
[33] Mitra, "Mitra," https://mitrarobot.com//, 2022, [Online; accessed 11-August-2022].
[34] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 2014, pp. 1061–1068.
[35] I. Poggi, C. Pelachaud, F. d. Rosis, V. Carofiglio, and B. D. Carolis, "Greta. a believable embodied conversational agent," in *Multimodal intelligent information presentation*. Springer, 2005, pp. 3–25.
[36] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robotics Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012.
[37] M. Fares, C. Pelachaud, and N. Obin, "Multimodal generation of upper-facial and head gestures with a transformer network using speech and text," *arXiv preprint arXiv:2110.04527*, 2021.
[38] C. Viegas, M. Inan, L. Quandt, and M. Alikhani, "Including facial expressions in contextual embeddings for sign language generation," *arXiv preprint arXiv:2202.05383*, 2022.
[39] J. Zelinka and J. Kanis, "Neural sign language synthesis: Words are our glosses," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 3395–3403.