Smile and Laugh Dynamics in Naturalistic Dyadic Interactions: Intensity Levels, Sequences and Roles

Kevin El Haddad
Numediart Institute - UMONS
Mons, Belgium
kevin.elhaddad@umons.ac.be

Sandeep Nallan Chakravarthula
University of Southern California
Los Angeles, United States
nallanch@usc.edu

James Kennedy
Disney Research Los Angeles
Glendale, United States
james.r.kennedy@disney.com

ABSTRACT

Smiles and laughs have been the subject of many studies over the past decades, due to their frequent occurrence in interactions, as well as their social and emotional functions in dyadic conversations. In this paper we push forward previous work by providing a first study on the influence one interacting partner's smiles and laughs have on their interlocutor's, taking into account these expressions' intensities. Our second contribution is a study on the patterns of laugh and smile sequences during the dialogs, again taking the intensity into account. Finally, we discuss the effect of the interlocutor's role on smiling and laughing. In order to achieve this, we use a database of naturalistic dyadic conversations which was collected and annotated for the purpose of this study. The details of the collection and annotation are also reported here to enable reproduction.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *User models*; *User studies.*

KEYWORDS

smiles, laughs, mimicry, conversation, dyadic interaction

ACM Reference Format:
Kevin El Haddad, Sandeep Nallan Chakravarthula, and James Kennedy. 2019. Smile and Laugh Dynamics in Naturalistic Dyadic Interactions: Intensity Levels, Sequences and Roles. In *2019 International Conference on Multimodal Interaction (ICMI '19), October 14-18, 2019, Suzhou, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3340555.3353764

ICMI '19, October 14-18, 2019, Suzhou, China

ACM ISBN 978-1-4503-6860-5/19/05.
https://doi.org/10.1145/3340555.3353764

1 INTRODUCTION

This work is part of a project that aims to build nonverbal expression-aware interactive virtual agents. In this framework, we focus on smiles and laughs (S&L) in particular. An important aspect of S&L is how they influence each other in dyadic interactions. Understanding this would help generation, prediction and recognition systems by taking into account user S&L in creating agent behavior. To our knowledge, this aspect has not been quantified in previous studies.

The expressions used during an interaction are influenced by context in the broader sense of the term (the emotional states of the participants, their own social and cultural background as well as their interlocutors', etc.). We therefore consider two main conversational behaviors to quantify S&L dynamics. First, mimicry is linked to the effect a participant has on their interlocutor(s). It reflects the inter-participant influence. Second, and in contrast, the study of sequences reflects the intra-participant influence since it represents the dependency of the current expression on the previous ones produced by a single participant.

For mimicry, it has been shown that smiles and laughs are contagious and induce an interlocutor to smile and laugh [9, 13]. However, intensity has not been taken into account, which is an important factor to consider as it is linked to the context and expression's functionality [11, 12]. In [14], the authors report that laughs cause recipients to smile or laugh, but these S&L were not in the context of an interaction. Another study focuses on the presence of synchronicity of S&L in humorous contexts [6], but does not quantify the synchronicity or mimicry between these expressions.

The emergence of sequences and patterns for a specific subject in a specific context is under-studied in existing literature. The recent push toward deep learning models for generating such patterns works well for similar tasks, but the black-box nature of the models limits understanding of the underlying phenomena. In this direction, Haakana [7] observed successions of S&L between interlocutors, but the study does not concern temporal successions of S&L for a specific interlocutor and does not provide quantitative results. To build an S&L-aware virtual agent, we need to better understand the influences of smiles and laughs on smiles or laughs, taking into account parameters characterizing

the S&L and the interaction, such as the intensity and the participants' conversational roles. We contribute by:

(1) Collecting and annotating a database of naturalistic dyadic conversations, with evaluation of the annotation protocol for reproduction.
(2) Quantifying the S&L influence between speakers and listeners by calculating the mimicry, taking the intensity into account.
(3) Statistically studying S&L sequences in time, taking into account the intensity and conversational role.

The authors in [6] consider that "smiling is intended as a continuum encompassing 'laughing smile' ". Other work has presented arguments in favor of the existence of a smile-laugh continuum, or contrarily, S&L being two different expressions [18]. In this study we also provide arguments that tend to favor a S&L continuum for the context of this data.

The results from this study can be used as a basis for developing HAI systems with S&L expressions. They can also be used as a baseline to objectively pre-evaluate the performance of machine learning-based systems before undergoing costly and time consuming subjective evaluations (such as in [4]).

2 DATABASE

A naturalistic dyadic interaction dataset consisting of 46 interactions was collected and annotated. Participants in the dataset were recruited from a business campus. A subset of the data, 15 interactions, was randomly selected for use in this work. Each dyad is a unique pair and the sessions are an average of 14 mins long (*SD* 2m-07).

Each interaction follows the same structure, using questions as prompts designed to elicit a variety of emotions and expressions in conversation. After signing consent forms, the participants are taken into the recording room and are fitted with a lavalier microphone. The experimenter delays providing the first prompt to give the participants the chance to get acquainted. Screens in the center of the room are used to display the question prompts. The experimenter monitors the conversation and moves to the next prompt after approximately 2 minutes, or sooner if the conversation has stopped. This process repeats for around 10 minutes. The prompts can be seen in the supplementary material.

The main goal of this dataset was to obtain multimodal conversation data with as much variability in emotions as possible. Inspired from previous work [8], the prompts were chosen to elicit this variety of emotional expressions by suggesting subjects related to emotional memories (negative and positive ones).

The physical setup can be seen in Fig. 1. An Intel RealSense camera is directed toward each participant, capturing RGB images. Each participant has a lavalier microphone, routed through a noise gate to remove as much overlapping interlocutor speech as possible. Video conferencing software is used with screen sharing so that the experimenter can monitor the interaction and update the prompts.

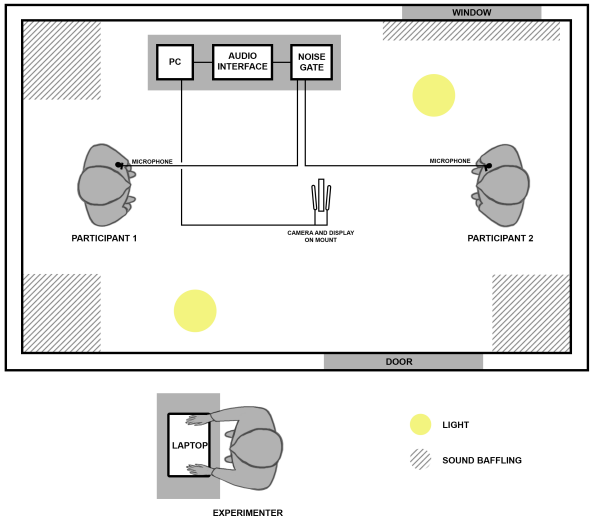

Figure 1: Recording setup. Figure not to scale

Annotation

Four annotators were trained to annotate phenomena in the dataset using ELAN [19]. The training guide can be found in the supplementary material, but is summarized below.

Each annotator was given an entire session to annotate (including both interlocutors). For the turn roles, the annotators were asked to segment each video in "speaker" (spk) and "listener" (lsn) segments. A "speaker" is the interlocutor who is uttering a full sentence or word (short or long), while the "listener" is the interlocutor to whom the speaker is talking. Lsn is thus the one giving feedback to spk or waiting for a response or, more generally, a message from their interlocutor. The segments start when the participant starts speaking and finish at the end of the utterance or any nonverbal expression (facial expression or body movement) co-occurring or directly following the utterance.

For smiles, the annotators were asked to focus visually on facial expressions using the activation of the zygomaticus, pulling the lip corners, lifting the cheeks, and activation of the orbicularis oculi (around the eyes). Smiles can also be perceived audibly while co-occurring with speech [1, 16]. An intensity is given to each smile segment and is used to delineate the segments. A smile segment starts at the frame at which the audio or visual expression starts, and stops either at the beginning of a smile of a different level, or at the end of the transition to a non-smile expression.

Laughs differ from smiles because they are expressed through exhalation/inhalation sequences of voiced and unvoiced sounds, and by body or periodic head movements. A laugh segment begins when the expression is perceived and ends either when no more expression is perceived, or at the end of an inhalation sound. A subsequent laugh is annotated as a new segment.

The annotators were not trained in Facial Action Coding System [3], so they were asked to label each segment with an intensity based on their perception of the expression without considering the emotion, affect or context that might be underneath it. The levels in increasing intensity order are low, medium and high. Classifying some expressions into smiles can be confusing due to low activation of the muscles involved and co-occurrence with another expression. Based on this, and previous smile annotation experiences, a fourth smile level was added below "low". This "subtle" level is intended to reduce the time deciding the nature of these expressions and to avoid omitting some expressions.

While laughs can contain smiles, we decided that S&L segments cannot temporally co-occur. This is to facilitate analysis and comparison by completely separating S&L.

Smiles and Laughs Inter-Rater Agreement

To estimate the inter-rater agreement, six randomly chosen files were annotated by a second coder (i.e., an annotation team member who had not coded that file). Cohen's kappa was calculated with, and without taking intensity into account. We obtain good results for segment overlap without considering intensity: κ=0.753 for laughs and 0.581 for smiles. Omitting the "subtle" smile segments (which had less strict annotation criteria), κ increases to 0.672. Taking the intensity into account: κ=0.585 for the laugh segments, 0.413 for the smiles with, and 0.480 without, the subtle class. Cohen's kappa on the turn roles also shows good agreement: κ=0.794.

3 MIMICRY

We first look at the general amount of S&L per role. Table 1 shows the mean duration of S&L per role and intensity. Comparing the corresponding levels of the spk and the lsn, we can see that the spk have, on average, shorter laughs and longer smiles than lsn. Only the smiles (except subtle) and the low laughs showed statistically significant differences (Student's t-test with 5% significance level), but these results suggest a possible influence between spk and lsn's S&L durations. The effect of the roles on the durations is unclear and could be investigated in the future.

Mimicry can be described as the replication or mirroring of one's expression by the interlocutor. To quantify mimicry, we applied a method similar to the one described in [5], which was also used in other work [2, 17]. For event B to mimic event A, B must occur after A's start and can continue until A's stop within a margin ΔT. In order to avoid double counting mimicry, B should stop before the next A starts. So, to count an event as mimicry the following must apply:

	SM				LGH		
	sub.	**low**	**med.**	**high**	**low**	**med.**	**high**
SPK	2.36	2.49	2.55	1.48	0.97	1.33	1.49
LSN	2.36	2.03	1.85	1.38	1.15	1.42	1.78

Table 1: Mean S&L durations in seconds by level and role.

$$T_{start}(A_i) < T_{start}(B_i) \quad (1)$$

$$T_{start}(B_i) < min\{T_{stop}(A_i) + \Delta T, T_{start}(A_{(i+1)})\} \quad (2)$$

Where B_i and A_i are respectively the i_{th} event in sequences of events and T_{start} and T_{stop} are respectively the starting and stopping times of an event. Here ΔT = 0 (0.5, 1, 1.5 and 2 seconds were also tested with similar results).

To quantify mimicry and compare it across the entire dataset, we use the probability that an expression B_i mimics A. We therefore calculate:

$$\frac{\sum_{n=0}^{N} mBA}{\sum_{n=0}^{M} B_n} \quad (3)$$

Which represents B mimicking A (mBA) over all occurrences of B.

By definition, an event is mimicry when the same expression is replicated. For the purpose of this study, and to analyse the influence of two expressions on each other (smiles and laughs), the definition is extended to copying different expressions as well, i.e., smiles mimicking laughs and vice versa. We thus calculate the mimicry of each expression at a given intensity on another, and spk on lsn and vice versa. The mean value is then calculated for turn role segments[1]. The results are shown in Fig. 2.

We first observe that, in general, laughs are mostly mimicked by laughs and not by smiles (Fig. 2-**d** vs **h**) while smiles are mimicked by both (Fig. 2-**a, b, e** and **f**). Then, when laughs mimic smiles, the levels of laughs are lower (Fig. 2-**b** and **f**), when smiles mimic laughs the smile levels are higher (Fig. 2-**c** and **g**). We can see this in most cases except when, in the lsn mimics spk condition, higher levels of smiles are mimicked by laughs (Fig. 2-**f**). In this case, high levels of laughs are most likely to happen. For smiles mimicking smiles, the levels mimicked have similar values (Fig. 2-**a** and **e**). Laughs mimic laughs when spk mimics lsn, however, when lsn mimics spk, laugh mimicry levels are lower.

Lower levels of laughs mimicking smiles and higher levels of smiles mimicking laughs are in favor of the smile-laugh continuum theory mentioned earlier with smiles being on

[1]A different implementation is used here, but this is done in a similar manner to publicly available tools such as: https://github.com/kelhad00/CBA-toolkit

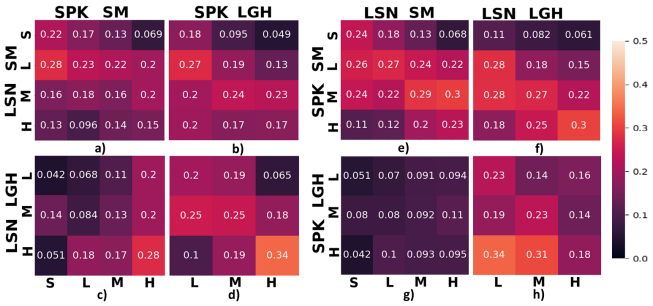

Figure 2: Mean mimicry probability heatmap. The horizontal label indicates the role of the mimicking participant, e.g., a) is SPK SM mimicking LSN SM. In each heatmap, the columns indicate the mimicking expression intensity, while the rows indicate the mimicked expression intensity. A higher cell value corresponds to a higher probability.

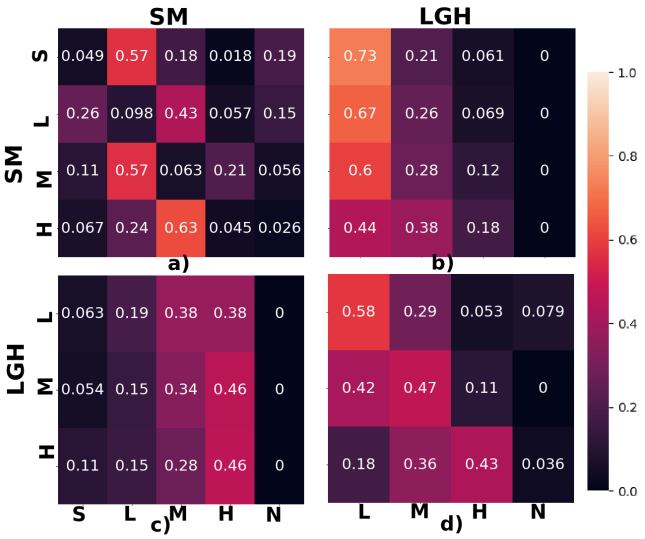

Figure 3: Probabilities that smiles (SM) or laughs (LGH) (columns) follow smiles or laughs (rows) at different intensities. L: low, M: medium, H: high, N: no S&L and, for smiles S: subtle. In the heatmaps, the S&L in the columns follow the ones in the rows.

the low arousal side and laughs on the high side of a common arousal level scale for both S&L.

The definition of mimicry implies that when S&L produce the same expression, the intensities should also be mimicked. This is the case here except for laughs mimicking laughs in the lsn mimic spk case. The relationship between laughs being volitional/fake or spontaneous/real and arousal was studied in [10]. Fake laughs are inversely proportional to arousal while real ones are proportional to it. Therefore, a possible interpretation for this exception is that spk is producing an utterance, expecting a reaction from lsn, so spk will mimic the laughs with real laughs and with a similar level as the lsn's. However, when lsn mimics spk, lsn could be producing fake laughs, and so laughs of lower levels.

4 SMILE AND LAUGH SEQUENCES

In this section we study the temporal sequence patterns of S&L produced during an entire interaction (i.e., one's own S&L sequence pattern, not considering their interlocutor's). For this, we consider the S&L directly following any smile or laugh. We first note that, in general, laughs and smiles are mostly followed by smiles (83.5% of laughs and 72.7% of smiles are followed by smiles). Fig. 3 shows the probabilities that S&L at different levels directly follow a specific expression in time (within a 500ms margin of error).

We can see that laughs are rarely followed by no S&L and mostly followed by smiles of high levels (Fig. 3-**c** and **b**). Also, when the smiles are followed by laughs, the laughs are mostly of lower levels. This is due to the rare occurrence of laughs of high levels compared to lower levels (86 high, 205 medium, 359 low). But we can observe with high probability, laughs with higher levels follow smiles with higher levels. When smiles follow smiles, we can see that the lower levels (subtle and low) are mostly followed by the levels directly above them (low and medium respectively) and the higher levels (medium and high) by the ones directly below them (medium and low respectively). When laughs follow laughs, the successive laughs have similar levels.

The smile-laugh continuum suggests that S&L can be represented on the same scale. Previous work investigated the existence of this continuum without a clear conclusion [15, 18]. Besides describing the S&L succession patterns, our observations here are in favor of this continuum. Indeed they suggest that S&L (especially laughs) rarely occur without any surrounding S&L, and that a relationship exists between the S&L intensity levels when these form sequences. In all cases, the most probable levels following other levels are the closest ones. There is thus, in a sequence of S&L, no sudden variation across levels but rather gradual variation.

5 CONCLUSION AND FUTURE WORK

In this paper, a database of dyadic interactions was collected and annotated. From this, we presented studies concerning the influences of smiles and laughs between speakers and listeners and S&L time sequence patterns for dyadic interaction participants. These help us to better understand the S&L dynamics in such contexts and will be used to build better models for S&L-aware behavior. The findings of this work contribute to the foundation of future S&L-aware agents.

REFERENCES

[1] Véronique Aubergé and Marie Cathiard. 2003. Can we hear the prosody of smile? *Speech Communication* 40, 1-2 (2003), 87–97.

[2] Sanjay Bilakhia, Stavros Petridis, Anton Nijholt, and Maja Pantic. 2015. The MAHNOB Mimicry Database: A database of naturalistic human interactions. *Pattern Recognition Letters* 66 (2015), 52–61.

[3] Paul Ekman. 1977. Facial Action Coding System. (1977).

[4] Kevin El Haddad, Hüseyin Çakmak, Emer Gilmartin, Stéphane Dupont, and Thierry Dutoit. 2016. Towards a Listening Agent: A System Generating Audiovisual Laughs and Smiles to Show Interest. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI 2016)*. ACM, New York, NY, USA, 248–255. https://doi.org/10.1145/2993148.2993182

[5] Sebastian Feese, Bert Arnrich, Gerhard Tröster, Bertolt Meyer, and Klaus Jonas. 2012. Quantifying behavioral mimicry by automatic detection of nonverbal cues from body motion. In *Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*. IEEE, 520–525.

[6] Elisa Gironzetti, Lucy Pickering, Meichan Huang, Ying Zhang, Shigehito Menjo, and Salvatore Attardo. 2016. Smiling synchronicity and gaze patterns in dyadic humorous conversations. *Humor* 29, 2 (2016), 301–324.

[7] Markku Haakana. 2010. Laughter and smiling: Notes on co-occurrences. *Journal of Pragmatics* 42, 6 (2010), 1499–1512.

[8] Louise Heron, Jaebok Kim, Minha Lee, Kevin El Haddad, Stéphane Dupont, Thierry Dutoit, and Khiet Truong. 2018. A Dyadic Conversation Dataset on Moral Emotions. In *Proceedings of the 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*. 687–691. https://doi.org/10.1109/FG.2018.00108

[9] Ursula Hess and Patrick Bourgeois. 2010. You smile–I smile: Emotion expression in social interaction. *Biological psychology* 84, 3 (2010), 514–520.

[10] Nadine Lavan, Sophie K Scott, and Carolyn McGettigan. 2016. Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter. *Journal of Nonverbal Behavior* 40, 2 (2016), 133–149.

[11] JS Lockard, CE Fahrenbruch, JL Smith, and CJ Morgan. 1977. Smiling and laughter: Different phyletic origins? *Bulletin of the Psychonomic Society* 10, 3 (1977), 183–186.

[12] Gary McKeown and Will Curran. 2015. The relationship between laughter intensity and perceived humour. In *The 4th Interdisciplinary Workshop on Laughter and other Non-Verbal Vocalisations in Speech*. 27–29.

[13] Costanza Navarretta. 2016. Mirroring Facial Expressions and Emotions in Dyadic Conversations. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 469–474.

[14] Robert R Provine. 1992. Contagious laughter: Laughter is a sufficient stimulus for laughs and smiles. *Bulletin of the Psychonomic Society* 30, 1 (1992), 1–4.

[15] Willibald Ruch and Paul Ekman. 2001. The expressive pattern of laughter. In *Emotion, qualia and consciousness*, A. Kaszniak (Ed.). World Scientific Publishers, Tokyo, 426–443.

[16] Marc Schröder, Véronique Aubergé, and Marie-Agnes Cathiard. 1998. Can we hear smile?. In *Proceedings of the Fifth International Conference on Spoken Language Processing*.

[17] Juan R Terven, Bogdan Raducanu, María Elena Meza-de Luna, and Joaquín Salas. 2016. Head-gestures mirroring detection in dyadic social interactions with computer vision-based wearable devices. *Neurocomputing* 175 (2016), 866–876.

[18] Jürgen Trouvain. 2001. Phonetic aspects of "speech laughs". In *Oralité et Gestualité: Actes du colloque ORAGE, Aix-en-Provence. Paris: L'Harmattan*. 634–639.

[19] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. 1556–1559.