Laughter and Other Non-Verbal Vocalisations Workshop 2020
5 October 2020, Bielefeld, Germany

Cross-Corpora Study of Smiles and Laughter Mimicry in Dyadic Interactions

Kevin El Haddad
numediart institute, UMONS
Mons, Belgium
kevin.elhaddad@umons.ac.be

Thierry Dutoit
numediart institute, UMONS
Mons, Belgium
thierry.dutoit@umons.ac.be

Abstract

In this paper, we present preliminary results of our ongoing work on cross-corpora analyses of smiles and laughter mimicry. For this, instead of recording new data, we leverage the ones produced and available. We analyze smiles and laughs mimicry in three different datasets and show results similar to our previous work. The data used here can be accessed at: https://doi.org/10.5281/zenodo.3820510.

1 Introduction

Smiles and Laughs (S&L) are important expressions to consider in dyadic interaction-related applications due to their frequency of occurrence and the broad range of functionality they have in such context. A plethora of work can be found on these two expressions in many different domains. It has been shown, on one side that S&L are contagious (Hess and Bourgeois, 2010; Navarretta, 2016) and that a mirroring or mimicry effect exists, and on another, that they play important roles in interactions (Lockard et al., 1977; McKeown and Curran, 2015). In our previous work S&L dynamics were studied on a dataset of dyadic interactions, among which the S&L mimicry between interlocutors (El Haddad et al., 2019). In that work, two main aspects of S&L dynamics were studied: mimicry of one interlocutor's expressions onto the other, and the potential influence of these expressions had on each other in a sequence for the same speaker. In both these aspects, parameters like the interlocutors' roles in the conversation (speaker or listener) and the S&L intensities were taken into account. That work was done on a single dataset of dyadic interaction recordings. In this paper we show first results on our attempt of reproducing our previous work on larger and more diverse data. Indeed, the study done in (El Haddad et al., 2019) was focused on a specific group of people and the data were manually annotated by four annotators. Replicating this work on datasets recorded in different contexts, for different purposes and with participants of different backgrounds would help better understand the dynamics of S&L in interactions and reduce the annotations biases like the annotator's subjectivity from the analyses.

2 Data used

All the datasets considered here contain dyadic conversations, but each were recorded for different purposes, with participants of different backgrounds, in different contexts and environments. We focused on three main datasets. In the Cardiff Conversation Database (CCDB) (Aubrey et al., 2013), the interactions were in English and the participants were presumably, in majority, of British background and so, native English speakers. They were free to discuss any topic even though some general topics were sometimes suggested to them. The IFA Dialog Video Corpus (IFADV) (Van Son et al., 2008) interactions were in Dutch with scripted and freely spoken data. The participants were presumably mostly of Dutch background and so, are native Dutch speakers. For the Naturalistic Dyadic Conversation on Moral Emotions (NDC-ME) database[1] (El Haddad et al., 2018a,b), the interactions were in English with participants of different backgrounds. These latter asked each other pre-assigned questions in turns. Most of the participants in this dataset are not native English speakers. So the main differences between these datasets reside in the languages used, the participants' backgrounds and the topics discussed but also notably the angle of the recording cameras, the recording environment and the quality of the data recorded differ.

[1] Please contact the authors for access to NDC-ME.

3 Annotation

The Roles (speaker/listener), the smiles and the laughs were annotated by a single annotator following the protocol described in (El Haddad et al., 2018b, 2019). At the moment of writing, the dataset was annotated only partially due to the time consuming nature of this process and the resources required. A session representing the entire interaction between both participant, we ended up with 7 sessions (14 participants) for CCDB, 8 for IFADV and 4 for NDC-ME. But only the first 1 or 2 minutes (depending on the session) were annotated for CCDB and IFADV while the entire sessions were annotated for NDC-ME. We obtained a total of 19 min of annotated data for CCDB, 26 min for IFADV and 95 min for NDC-ME (the sessions were longer than the others). The S&L intensities were also annotated but were not taken into account in this study

4 Mimicry Definition

By definition, one's expression is being mimicked when it is replicated by the interlocutor. Mimicry was previously studies in the literature. We based our implementation of it on the previous works (Feese et al., 2012; Terven et al., 2016)[2]. For event B to mimic event A, B must begin 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:

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

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

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 $\Delta T = 0$ (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:

$$\frac{\sum_{n=0}^{N} mBA}{\sum_{n=0}^{M} B_n} \tag{3}$$

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

To be precise, mimicry is usually only considered for the same expressions. But for the sake of this study, this will be extended to also included smiles mimicking laughs and vice-versa.

[2]Please refer to the implementation in the CBA toolkit

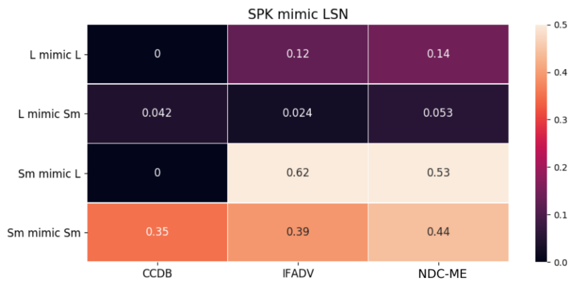

Figure 1: Mean mimicry probabilities for Speaker (SPK) mimicking Listener (LSN)'s smiles (Sm) and Laughs (L)

5 Results

Mimicry was thus calculated for each speaker/listener segments, per expression and for each session. We thus obtain one value representing the probability of a mimicry event to occur for each pair of videos annotated. The mean value of was calculated to represent mimicry for each pair of expressions and dataset considered. The results are shown below. They show the mean mimicry values for the smiles (Sm) and laughs (L) when the speaker (SPK) mimics the listener (LSN) (Fig. 1) and vice-versa (Fig. 2).

The observations that will follow are obviously not representative of the S&L dynamics in general in dyadic conversations considering the imbalance in the data used here and the fact that it was annotated by a single person. Also other parameters such as the intensity are important to consider as suggested by our previous study (El Haddad et al., 2019). Although these were annotated, they were not considered for this study but will be the subject of future work. Nevertheless, these observations give interesting first insights.

The main common points between these results and the ones in (El Haddad et al., 2019)), are the following: i) smiles seem to mimic smiles in relatively high probabilities and in all cases; ii) laughs mimicking smiles seem to have low probabilities in all cases; iii) smiles mimicking laughs seem to have a rather high probability; iv) laughs mimicking laughs' probability is rather high when LSN mimic SPK.

A main difference to note is that, even though laughs mimicking laughs has a rather low probability for all three datasets studied here, it seemed

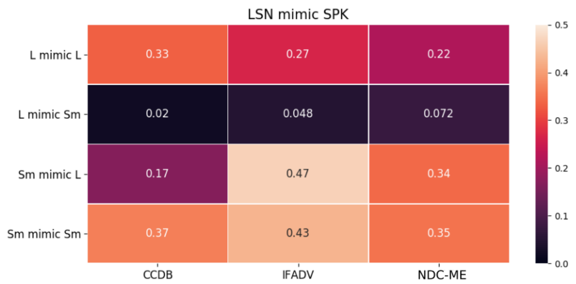

Figure 2: Mean mimicry probabilities for Listener (LSN) mimicking Speaker (SPK)'s smiles (Sm) and Laughs (L)

to have a high probability in our previous work. This might be due to the amount of data studied or might be related to the context, topic or participants behaviors. Further work should be made for a clear conclusion. These results show that the observations made in our previous work might be generalized to broader interaction contexts.

6 Conclusion

In this paper we presented our ongoing work on cross-corpora nonverbal expressions dynamics. Here, we compared S&L mimicry across datasets and observed several similarities with our previous work. These first results are encouraging for future work which aim at improving the way we interact with virtual agents. Indeed, a better understanding of nonverbal expression dynamics will serve as a benchmark for future evaluations and a reference for building and debugging data-driven systems.

Finally, with this work, we introduce our attempt to build a large database formed of the data produced by, and already available for the community. This is motivated mainly by the need of more data to improve intelligent systems in Human-Agent Interaction applications. In the rise of deep learning and data hungry systems, the quantity of data available for specific tasks is important. Although, there exist several datasets available containing annotated data of smile or laughter, or even both, they are individually either not sufficient or not adequate to train systems for specific tasks. Some of the reasons for this are: i) the data available do not have a homogeneous format (modality, recording setup, etc.), ii) the content can vary greatly because of the context of recording for instance, iii) the annotations, if available, were made with different annotation protocols and tools and so, are usually different from one dataset to another.

References

Andrew J Aubrey, David Marshall, Paul L Rosin, Jason Vendeventer, Douglas W Cunningham, and Christian Wallraven. 2013. Cardiff conversation database (ccdb): A database of natural dyadic conversations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 277–282.

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, page 259–263. ACM.

Kevin El Haddad, Yara Rizk, Louise Heron, Nadine Hajj, Yong Zhao, Jaebok Kim, Trung Ngô Trng, Minha Lee, Marwan Doumit, Payton Lin, et al. 2018a. *End-to-End Listening Agent for Audiovisual Emotional and Naturalistic Interactions*, volume 10.

Kevin El Haddad, Noé Tits, and Thierry Dutoit. 2018b. Annotating nonverbal conversation expressions in interaction datasets. In *5th Interdisciplinary Workshop on Laughter and Other Non-Verbal Vocalisations in Speech*, pages 54–47, Paris, France.

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*, pages 520–525. IEEE.

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

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

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*, pages 27–29.

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)*, pages 469–474.

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*, 175:866–876.

RJJH Van Son, Wieneke Wesseling, Eric Sanders, Henk van den Heuvel, et al. 2008. The ifadv corpus: A free dialog video corpus.