Communicative Signals and Social Contextual Factors in Multimodal Affect Recognition

Michel-Pierre Jansen
University of Twente
Human Media Interaction
Enschede, Overijsel
m.jansen-1@utwente.nl

ABSTRACT

One research branch in Affective Computing focuses on using multimodal 'emotional' expressions (e.g. facial expressions or non-verbal vocalisations) to automatically detect emotions and affect experienced by persons. The field is increasingly interested in using contextual factors to better infer emotional expressions rather than solely relying on the emotional expressions by themselves. We are interested in expressions that occur in a social context. In our research we plan to investigate how we can; a) utilise communicative signals that are displayed during interactions to recognise social contextual factors that influence emotion expression and in turn b) predict/recognise what these emotion expressions are most likely communicating considering the context. To achieve this, we formulate three main research questions: I) How do communicative signals such as emotion expressions co-ordinate behaviours and knowledge between interlocutors in interactive settings?, II) Can we use behavioural cues during interactions to detect social contextual factors relevant for interpreting affect? and III) Can we use social contextual factors and communicative signals to predict what emotion experience is linked to an emotion expression?

CCS CONCEPTS

• **Human-centered computing**;

KEYWORDS

Affective Computing, Automatic Affect Detection, Multimodal Emotion Recognition, Emotion Expressions, Social Contextual Factors, Communicative Signals, Interpersonal Dynamics

1 INTRODUCTION

Affective Computing (AC) has come a long way since the founding of the field [19]. One of the goals of the field is to approach accurate automatic affect detection. This could have several useful implications including providing systems with the tools to better understand and communicate with humans, and provide affect researchers with additional tools to help study the different components of affect [6]. To attain this goal, researchers in the AC field draw upon and integrate knowledge from several disciplines including computer science, engineering, philosophy, psychology and neuroscience among many others.

Until recently there was a focus on using behavioural expressions from the expressor to categorise emotions that are displayed by persons. However, there seems to be increasing evidence that there is great variance between persons and situations in how people express the emotions they indicate to experience [2, 3, 8, 15, 20]. This poses a challenge for attaining accurate automatic affect detection across person and situation. One alternative approach is to follow a more constructionist approach for affect detection. According to this approach, human observers are able to categorise emotions despite the variance in emotional expressions by using contextual factors surrounding the expressor to predict which emotion best represents the behavioural expression of interest. Some proponents of this approach point out that a similar process happens with the experience of an emotion by the expressor, where the expressor utilises contextual factors to categorise a general internal valenced feeling [4].

There is a surge of interest within the Affective Computing domain, and more generally emotion research, to explore

how context can be mapped and leveraged for affect detection [1, 11, 14]. However, relevant context remains difficult to identify or apply in affect detection.

One relevant source of context for emotion and affect detection could be the social context, since emotion expressions seem to occur more consistently and regularly in settings that have some form of social interaction between people [8, 9, 15]. For example, we often only smile when we are interacting with another person [9]. In this light, emotion expressions could be thought of more as communicative signals rather than solely as expressions of internal states that automatically show up [2, 15]. It has been argued by various authors that emotional expressions have several communicative functions [21] that extend beyond expressing an emotion.

In our research we plan to investigate how we can; **a)** use communicative signals in interactions to recognise social contextual factors that influence emotion expression and in turn **b)** predict/recognise what emotion expressions are most likely communicating in that context.

2 BACKGROUND AND RELATED WORK

There is quite some evidence that suggests that there is a great variety in how people express their emotions [2, 3, 8, 15, 20]. Some articles [8, 20] reviewed available studies on the coherence between emotion experiences and (facial) expressions of emotions to test how consistent people were in expressing specific emotions. Reisenzein, Studtmann and Horstmann [20] focused on results from laboratory studies, while Fernandez-Dols and Crivelli [8] reviewed results from more naturalistic studies. Both articles concluded that in general, with the exception of some positive emotions like amusement, coherence between specific emotions and facial expressions seemed low or not significant.

Fernandez-Dols and Crivelli [8] offer in their discussion section an explanation for the low coherence rates. They point out that there are several theories that argue that everyday facial expressions that are often recognised as emotional expressions can have other causes and functions, for example to co-ordinate behaviour or knowledge between interlocutors [10]. The authors then continue this line of thought and argue that smiles, one of the most common interactive facial behaviours with a high coherence to positive emotions [8, 20, 22], serves multiple communicative functions (for a recent overview and proposed theory we refer to [21]).

When we view these "affective" facial expressions as communicative functions to co-ordinate behaviours and opinions, we expect that behaviours align in some ways between interlocutors and that some form of interdependence is established just as with other communicative acts. With respect to this hypothesised interdependence some research

has been published in the Affective Computing community focusing on closely linked subject, for example [23] used a cross-recurrence based methodology to investigate emotional contagion as a form of emotion expression interdependence.

3 PROBLEM STATEMENT AND KEY RESEARCH QUESTIONS

Scope: One research branch in Affective Computing focuses on using multimodal 'emotional' expressions (e.g. facial expressions or non-verbal vocalisations) to automatically detect emotions and affect experienced by persons. However, contrary to what has long been believed, more and more evidence points toward a great variance in how people express emotions.

The field is increasingly interested in using contextual factors surrounding the expressor to better interpret emotional expressions rather than solely relying on the emotional expressions of the expressor. Some attempts to map or identify context relevant to emotions and affect have been made in fields related to Affective Computing [1, 4, 11, 12]. However, context relevant to emotions and affect remains hard to define, map and more importantly operationalise for automatic affect detection.

As mentioned in the background section, it seems that so-called emotional expressions occur often in interactive settings. Quite a few researchers on (facial) expressions and emotion researchers view these emotional expressions rather as communicative signals that can carry communicative functions than expressions that are there solely for the purpose of expressing an inner emotional experience. It has been suggested that these communicative signals help people co-ordinate behaviours and knowledge. This forms an interdependence between the interlocutors when communicating and expressing emotions.

The social context therefore is an important source of context factors that can help Affective Computing researchers better understand, infer and detect affect. Social contextual factors are in general not yet explored much in our field. Aviezer, Ensenberg and Hassin [1] points out that an important contextual cue to interpret an emotional expression is how other interlocutors react to and interact with the expressor of this emotion. Our research aim is to explore how social contextual factors and communicative signals from interlocutors over the span of an interaction can help to predict and infer what specific emotion expressions from individual interlocutors mean.

- **Key research question 1:** How do communicative signals such as emotion expressions co-ordinate behaviours and knowledge between interlocutors in interactive settings?

- **Key research question 2:** Can we use behavioural cues during interactions to detect social contextual factors relevant for interpreting affect?
- **Key research question 3:** Can we use social contextual factors and communicative signals to predict what emotion experience is linked to an emotion expression?

4 RESEARCH PLAN: METHODS AND PROGRESS

RQ1: How do communicative signals such as emotion expressions co-ordinate behaviours and knowledge between interlocutors in interactive settings?

How does laughter co-ordinate perceived humour between participants? Smiles have been recognised as one of the most common emotion expressions that is also used as a communicative signal [8, 20]. For example Scherer and Ceschi [22] studied its relationship to perceived and experienced humour and found a positive correlation. Another often used signal in social interactions is laughter, which is a multimodal expression. Similar to smiling, laughter is associated with humour and can serve multiple functions. Laughter therefore is ideal as a first step to understand how multimodal emotional expressions co-ordinate behaviour and knowledge between interlocutors.

In this study we plan to investigate how several features of laughter signals co-ordinate the way people predict how funny they are perceived. A possible feature of interest is the intensity/pitch of laughs and responsive laughs, since some research [16] indicates that laughter intensity is positively linked to laughter being perceived as humorous laughter. Other laughter specific features we plan to investigate are the total number, average duration of laughs as well as synchronicity between laughter among interlocutors. We are also interested in other, non-laughter specific contextual features that could influence the way people predict how humorous they are perceived. These include gender of the interlocutors, type of task (e.g. structured vs non-structured) and lexical features.
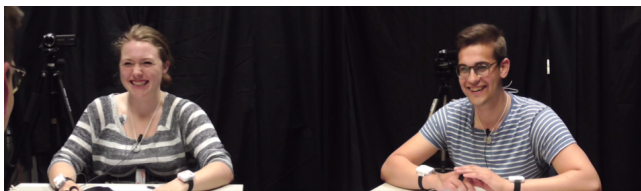

Figure 1: A snapshot of one of the sessions in our multimodal laughter in interaction database (MULAI)

Progress so far. Last year we constructed the MULAI database [13] in order to research how emotional expressions, in this case laughter, co-ordinate behaviours and knowledge

between interlocutors during interactions. The database contains synchronised video, audio and physiological data of 16 pairs having task-based interactions (a survival task, a relatively structured make-the-other-laugh challenge, and structured joke-telling rounds). For each round in the task-based interactions, both interlocutors of the pair filled in how humorous they perceived themselves and how humorous they perceived the other during that specific round. The latter two tasks have recently been annotated on laughter-bout level including in-breaths and out-breaths directly before and after these bouts. Preliminary results show that there is a general consensus between participants about how funny they find themselves and how funny they are perceived. Interestingly, this strongly varies with some of the other tasks. In the following months we hope to continue the annotation process as well as answer the primary research question.

RQ2: Can we use behavioural cues during interactions to detect social contextual factors relevant for interpreting affect?

Within this research question we hope to explore if we can use relatively "simple" behavioural cues that happen in interaction settings to predict or recognise social contextual factors. Recognising and knowing contextual factors could be helpful to later predict/categorise the meaning/intention of a more complex emotional expression.

"Simple" behavioural cues could include back-channels, eye movements, laughter, smiles, pitch of voice, hand gestures for example. These communicative expressions should be fairly easy to categorise/annotate and could possibly be outsourced. Crowd-sourcing techniques could be a plausible alternative if the manual annotation of collected materials is to labour intensive [17]. In addition it would be interesting to see how interlocutors respond to each other on these kind of expressions as a predictor for both the meaning of the expression and the social context.

A social contextual category of factors we are considering for our research is the relationship, closeness and familiarity between interlocutors. Some researchers point out that functional relationships between interlocutors are an important influence on interpersonal communication and emotion expression [1, 5]. Other contextual categories could be related to the kind of social interaction happening, for example the topic and function of interaction. Finally, we are interested in social contextual factors that are related to interlocutor characteristics such as the age, cultural orientation and genders and their specific combinations during interactions.

Progress so far. Recently we identified over 18 emotion annotated social interaction databases and reviewed several contextual factors available within them [7]. This review can be a starting point to either identify a suitable database for

our purposes or taking lessons from them and creating a new database to answer the primary research questions.

RQ3: Can we use social contextual factors and communicative signals to predict what emotion experience is linked to an emotion expression?

With this question we aim to explore if we can use social contextual factors and behavioural cues throughout the interaction to interpret what emotion experience is linked to an emotion expression at a specific time interval. We will not only focus on what happens at the moment of emotion expression, but also use the dynamics that happened earlier in the conversation. Since the full interaction can offer a lot of data about both interlocutors and how interlocutors respond to each other, we plan to study if we can leverage this in the interpretation of emotion expressions. We plan to use social contextual factors and behavioural cues that are also used for answering the second research question.

Examples of emotion expressions where social context could play a vital role in predicting its meaning are laughter or tears. Depending on the social context and communicative signals during the interaction, laughter could be interpreted as schadenfreude or humorous laughter whereas tears could be joy or sadness.

Can we use social contextual factors and behavioural cues to predict what emotion experience and expression will happen? If we are able to leverage social contextual factors and behavioural cues for affect detection, it might be interesting to actually see if we could use the same variables for predicting if (and when) interlocutors are going to express affect. We plan to explore possibilities with this idea in mind in the near future.

5   CONTRIBUTIONS AND CHALLENGES

Our research aims to explore how emotion expressions during interactions can be interpreted by looking at general characteristics of the interaction (social contextual factors) in combination with what happens in the interaction (communicative signals of interlocutors and their responses). Some scientific contributions that could be drawn from answering the research questions posed in our research plan are:

- Highlighting the importance of social contextual factors to better understand and recognise emotional expressions rather than solely relying on the emotional expressions.
- Identifying a set of relevant social contextual factors that can contribute in the prediction and categorisation of emotion expressions.
- Identifying ways to detect social context through communicative signals.

- Contributing to more accurate affect detection by taking contextual factors into account

To date we have contributed by identifying possible sources of context from an human perceiver perspective as well as reviewing 18 emotion annotated social interactional databases on what kind of context is already available in these databases, this is in the form of a conference publication [7]. In addition we have recorded two interactive databases, one focused mainly on laughter as a social signal [13] and the other on emotions in elderly [18].

We expect that we will face some challenges with regards to our research plan. The doctoral consortium will be an excellent place to refine the research plan, discuss possible challenges and receive feedback from peers and experienced researchers. Some of the challenges and questions we anticipate include; 1) What emotional expressions, behavioural cues and social contextual factors would be best suited for our research, and how to obtain usable data. 2) How to annotate the data in meaningful and useful emotion- and behavioural segments, and to use an existing annotation format or design a new one?

With this research plan we hope to make contributions to the field of Affective Computing and emotion research in general. A doctoral consortium at the ICMI 2019 would offer the opportunity to benefit from the feedback of experienced researchers and peers.

ACKNOWLEDGMENTS

REFERENCES

[1] Hillel Aviezer, Noga Ensenberg, and Ran R. Hassin. 2017. The inherently contextualized nature of facial emotion perception. *Current Opinion in Psychology* 17 (2017), 47–54. https://doi.org/10.1016/j.copsyc.2017.06.006

[2] Lisa Feldman Barrett. 2011. Was Darwin Wrong About Emotional Expressions? *Current Directions in Psychological Science* 20, 6 (2011), 400–406. https://doi.org/10.1177/0963721411429125

[3] Lisa Feldman Barrett, Kristen A. Lindquist, and Maria Gendron. 2007. Language as context for the perception of emotion. Trends in cognitive sciences. 11, 8 (2007), 327–332.

[4] Lisa Feldman Barrett, Batja Mesquita, and Maria Gendron. 2011. Context in Emotion Perception. *Current Directions in Psychological Science* 20, 5 (oct 2011), 286–290. https://doi.org/10.1177/0963721411422522

[5] Emily A. Butler. 2011. Temporal interpersonal emotion systems: The "TIES" that form relationships. *Personality and Social Psychology Review* 15, 4 (2011), 367–393. https://doi.org/10.1177/1088868311411164

[6] Sidney D'Mello, Arvid Kappas, and Jonathan Gratch. 2018. The Affective Computing Approach to Affect Measurement. *Emotion Review* 10, 2 (2018), 174–183. https://doi.org/10.1177/1754073917696583

[7] Bernd Dudzik, Michel-pierre Jansen, Franziska Burger, Frank Kaptein, Joost Broekens, Dirk K J Heylen, Hayley Hung, Mark Neerincx, and Khiet P. Truong. 2019. Context for Emotion Perception in Audiovisual Databases for Automatic Affect Detection: A Survey. In *International*

Conference on Affective Computing and Intelligent Interaction (ACII), in press. IEEE, 8.

[8] José Miguel Fernández-Dols and Carlos Crivelli. 2013. Emotion and expression: Naturalistic studies. *Emotion Review* 5, 1 (2013), 24–29. https://doi.org/10.1177/1754073912457229

[9] José Miguel Fernández-Dols and María Angeles Ruiz-Belda. 1995. Are Smiles a Sign of Happiness? Gold Medal Winners at the Olympic Games. *Journal of Personality and Social Psychology* 69, 6 (1995), 1113–1119. https://doi.org/10.1037/0022-3514.69.6.1113

[10] Alan J. Fridlund. 1994. *Human facial expression: An evolutionary view.* Academic Press, Santa Barbara. 369 pages.

[11] Katharine H. Greenaway, Elise K. Kalokerinos, and Lisa A. Williams. 2018. Context is Everything (in Emotion Research). *Social and Personality Psychology Compass* 12, 6 (jun 2018), e12393. https://doi.org/10.1111/spc3.12393

[12] Ursula Hess and Shlomo Hareli. 2015. The influence of context on emotion recognition in humans. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 1–6. https://doi.org/10.1109/fg.2015.7284842

[13] Michel-pierre Jansen, Dirk K. J. Heylen, Khiet P. Truong, Gwenn Englebienne, and Deniece S Nazareth. 2018. The MULAI Corpus: Multimodal Recordings of Spontaneous Laughter in Dyadic Interaction. In *Proceedings of Laughter Workshop 2018*, Jonathan Ginzburg and Catherine Pelachaud (Eds.). Paris, 58–63.

[14] Ronak Kosti, Jose M Alvarez, Adria Recasens, and Agata Lapedriza. 2017. Emotion Recognition in Context. In *The IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1667–1675.

[15] Kristen A. Lindquist and Maria Gendron. 2013. What's in a word? Language constructs emotion perception. *Emotion Review* 5, 1 (2013), 66–71. https://doi.org/10.1177/1754073912451351

[16] Gary McKeown and Will Curran. 2015. The relationship between laughter intensity and perceived humour. In *Proceedings of the 4th Interdisciplinary Workshop on Laughter and Other Non-verbal Vocalisations in Speech*. 27–30.

[17] R. Morris and Daniel McDuff. 2015. Crowdsourcing Techniques for Affective Computing. In *Handbook of Affective Computing*, Rafael A. Calvo, Sidney K. D'mello, Jonathan Gratch, and Arvid Kappas (Eds.). Oxford University Press Oxford, New York, NY, 384–394. https://doi.org/10.1177/1354856507084420 arXiv:doi:10.1038/nature09304

[18] Deniece S Nazareth, Michel-pierre Jansen, Khiet P. Truong, Gerben Westerhof, and Dirk K. J. Heylen. 2019. MEMOA: Introducing the Multi-Modal Emotional Memories of Older Adults database. In *International Conference on Affective Computing and Intelligent Interaction (ACII), in press*. IEEE, 7.

[19] Rosalind W Picard. 2000. *Affective computing*. Technical Report. 16 pages. https://doi.org/10.1016/S1364-6613(98)01190-5

[20] Rainer Reisenzein, Markus Studtmann, and Gernot Horstmann. 2013. Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review* 5, 1 (2013), 16–23. https://doi.org/10.1177/1754073912457228

[21] Andrea Scarantino. 2017. How to Do Things with Emotional Expressions: The Theory of Affective Pragmatics. *Psychological Inquiry* 28, 2-3 (2017), 165–185. https://doi.org/10.1080/1047840X.2017.1328951

[22] Klaus R Scherer and Grazia Ceschi. 2000. Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Personality and Social Psychology Bulletin* 26, 3 (2000), 327–339.

[23] Giovanna Varni, Isabelle Hupont, Chloe Clavel, and Mohamed Chetouani. 2017. Computational Study of Primitive Emotional Contagion in Dyadic Interactions. *IEEE Transactions on Affective Computing* 3045, c (2017). https://doi.org/10.1109/TAFFC.2017.2778154